



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΤΙΤΛΟΣ

*Αντιστοίχιση εικόνων με χρήση αλγορίθμων εξαγωγής
χαρακτηριστικών*

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΦΩΤΙΑ ΒΑΣΙΛΕΙΟΥ



ΚΟΖΑΝΗ, ΙΟΥΝΙΟΣ 2013

Επιβλεπων : Ζυγκιρίδης Θεόδωρος
& Σπύρου Ευάγγελος

Περίληψη

Η αντιστοίχιση εικόνων με σκοπό την δημιουργία πανοράματος. Στην παρούσα εργασία έχουν χρησιμοποιηθεί αλγόριθμοι οι οποίοι βασίζονται στην εξαγωγή χαρακτηριστικών (SURF).

Αρχικά βρίσκονται τα σημεία ενδιαφέροντος για κάθε εικόνα με τη βοήθεια του αλγορίθμου SURF και στη συνέχεια έχοντας ολοκληρώσει αυτήν την διαδικασία για όλες τις εικόνες, προσπαθούμε να τις ενώσουμε. Για να γίνει αυτό πρέπει πρώτα να μελετήσουμε αν μπορούν να ενωθούν και η συγκεκριμένη διαδικασία γίνεται με το ταίριασμα των σημείων ενδιαφέροντος. Στη συνέχεια για να πετύχουμε υψηλής ακρίβειας αντιστοίχιση χρησιμοποιούμε τον αλγόριθμο RANSAC, με τη βοήθεια του οποίου υπολογίζουμε τον γεωμετρικό μετασχηματισμό ανάμεσα στις δύο εικόνες την λεγόμενη ομογραφία. Αν ο αριθμός των κοινών σημείων είναι επαρκής τότε γίνεται το λεγόμενο ταίριασμα τους.

Λέξεις κλειδιά

Εξαγωγή χαρακτηριστικών, σημεία ενδιαφέροντος, σημεία-κλειδιά, SIFT, SURF, ομογραφία, RANSAC, συρραφή εικόνων.

Ευχαριστίες

Για την παρούσα διπλωματική εργασία η οποία ολοκληρώθηκε την περίοδο 2012 – 2013 θέλω να ευχαριστήσω όσους συνέβαλαν για την εκπόνησή της. Αρχικά θέλω να ευχαριστήσω τους επιβλέποντες της διπλωματικής μου εργασίας , κ Σπύρου Ευάγγελο όπως επίσης και τον κ Ζυγκιρίδη Θεόδωρο για την πολύτιμη βοήθεια και καθοδήγησή τους. Ευχαριστώ ακόμη τους φίλους(ες) μου για την ηθική τους υποστήριξη. Τέλος είμαι ευγνώμων στην οικογένεια μου για την υποστήριξη και κατανόηση τους όλα αυτά τα χρόνια. Αφιερώνω αυτήν την εργασία στην μητέρα μου, τον πατέρα μου και τον αδερφό μου.

Περιεχόμενα

Κεφάλαιο 1.....	1
Εισαγωγή.....	1
1.1 Ψηφιακή εικόνα.....	1
1.2 Raster και Vector εικόνες.....	2
1.3 Διάφορες Εφαρμογές και πρότυπα Ψηφιακής εικόνας.....	3
1.4 Τι είναι το Matlab.....	7
Κεφάλαιο 2.....	9
Αλγόριθμοι SURF & SIFT.....	9
2.1 Εισαγωγή.....	9
2.2 Αλγόριθμος SIFT (Scale-invariant feature transform).....	11
2.2.1 Γενική επισκόπηση του αλγορίθμου.....	11
2.2.2 Βασικά στάδια αλγορίθμου.....	12
2.2.3 Ο αλγόριθμος SIFT και τα 4 βήματα.....	15
2.2.2 Εντοπισμός σημείων-κλειδιών.....	17
2.2.3 Ανάθεση προσανατολισμού.....	19
2.2.4 Περιγραφή των σημείων-κλειδιών.....	19
2.3 Αλγόριθμος SURF (Speeded-Up Robust Features).....	22
2.3.1 Τα σημεία ενδιαφέροντος.....	22
2.3.2 Προσδιορισμός των σημείων ενδιαφέροντος.....	24
2.3.3 Η μήτρα Hessian.....	26
2.3.4 Κατασκευή της κλίμακας χώρου.....	27
2.3.4 Περιγραφέας σημείων ενδιαφέροντος.....	29
Κεφάλαιο 3.....	32
Αλγόριθμος RANSAC (Random Sample Consensus).....	32
3.1 Εισαγωγή.....	32
3.2 Γενική επισκόπηση αλγορίθμου.....	32
3.3 Παράμετροι του αλγορίθμου.....	35
3.4 Πλεονεκτήματα και μειονεκτήματα.....	37
ΚΕΦΑΛΑΙΟ 4.....	38
Μεθοδολογία Εφαρμογής.....	38

4.1 Βασικές Λειτουργίες	38
4.2 Αντιστοιχισή σημείων.....	38
4.2.1 Υπολογισμός της ομογραφίας(homography)	39
4.2.2 Πιθανοτικό μοντέλο για την επαλήθευση της αντιστοίχισης εικόνας.	41
4.3 Συρραφή εικόνων	42
Κεφάλαιο 5.....	45
Παραδείγματα – Αποτελέσματα	45
5.1 Πρώτο παράδειγμα	45
5.2 Δεύτερο Παράδειγμα.....	46
Βιβλιογραφία.....	48

Κεφάλαιο 1.

Εισαγωγή

Μία εικόνα αξίζει όσο 1000 λέξεις, λέει μια παλιά κινέζικη παροιμία και ο καθένας μας γνωρίζει πόσο ευκολότερα κατανοήσιμη είναι μια εικόνα από την περιγραφή με κείμενο. Η παροιμία αυτή επαληθεύεται και από τους φυσιολόγους. Το ανθρώπινο οπτικό κανάλι μπορεί να αντληθθεί 30-40 Mbits πληροφορίας ανά δευτερόλεπτο. Θεωρώντας ότι ένας άνθρωπος μπορεί να κατανοήσει 600-1200 λέξεις γραπτού κειμένου το λεπτό και υποθέτοντας ότι κάθε λέξη αποτελείται από 4 γράμματα και κάθε γράμμα περιέχει 7 bits πληροφορίας, εύκολα αποδεικνύεται το αληθές της παροιμίας.

Η εικόνα είναι μία από τις σημαντικότερες πηγές πληροφορίας αν όχι η σημαντικότερη. Σε καθημερινή βάση ερχόμαστε σε επαφή με ψηφιακές εικόνες χάρη στην σημερινή τεχνολογία. Το πρόβλημα είναι πως ο περισσότερος κόσμος δεν έχει τις απαραίτητες γνώσεις έτσι ώστε να καταλάβει τι είναι ψηφιακή εικόνα και κατά συνέπεια να διαχειρίζεται με τον σωστό τρόπο τις ψηφιακές του εικόνες.

Τα τελευταία χρόνια έχουμε μια ραγδαία αύξηση του αριθμού των εικόνων που συναντάμε σε ψηφιακή μορφή. Η διάδοση της ψηφιακής φωτογραφικής μηχανής επέτρεψε στον καθένα να φωτογραφίζει γρήγορα και εύκολα τον κόσμο που τον περιβάλλει. Με ελάχιστο κόστος μπορεί πια να ανεβάσει τις συλλογές φωτογραφιών του στο διαδίκτυο και έτσι να τις μοιράζεται με φίλους και αγνώστους.

1.1 Ψηφιακή εικόνα

Αρχικά πριν αναφέρουμε οτιδήποτε για τη ψηφιακή εικόνα θα πρέπει να κατανοήσουμε μία έννοια που συναντάμε καθημερινά αλλά λογικά δεν γνωρίζουμε τι είναι στην πραγματικότητα. Ο λόγος για το Pixel ή εικονοστοιχείο. Το Pixel είναι το μικρότερο στοιχειώδες τμήμα από το οποίο αποτελείται μια εικόνα. Πολλά pixels μαζί κατανομημένα οριζόντια ή κάθετα σχηματίζουν κάθε εικόνα που βλέπουμε στην εικόνα του υπολογιστή μας.

Μία ακίνητη εικόνα είναι ένα σήμα $z = f(x,y)$ με δύο ανεξάρτητες μεταβλητές x και y . Ακόμη οι δύο γεωμετρικές διαστάσεις του υποστρώματος της εικόνας είναι το πλάτος και το μήκος της. Η εξαρτημένη μεταβλητή z αντιστοιχεί στην ένταση της φωτεινότητας της εικόνας.

Σε μία έγχρωμη εικόνα, η τιμή του σήματος σε ένα σημείο, είναι μια διατεταγμένη τριάδα θετικών αριθμών. Σε αυτήν την περίπτωση το χρωματικό μοντέλο είναι RGB (Red Green Blue) που σημαίνει ότι κάθε pixel έχει αντί για μία τιμή φωτεινότητας, τρεις τιμές, οι οποίες αντιστοιχούν στα τρία χρώματα. Στις ασπρόμαυρες εικόνες (grayscale) η ένταση είναι η ίδια και για τα τρία βασικά χρώματα.

Εικόνες όπου το πεδίο ορισμού και το πεδίο τιμών της συνάρτησης f δεν είναι όλοι οι πραγματικοί αριθμοί αλλά πεπερασμένοι ακέραιοι αριθμοί τότε η εικόνα ονομάζεται ψηφιακή. Σύμφωνα με τον [1] μία ψηφιακή εικόνα προκύπτει με τη διακριτοποίηση του πεδίου τιμών της αντίστοιχης αναλογικής εικόνας. Η διαδικασία μετατροπής μίας αναλογικής εικόνας σε ψηφιακή ονομάζεται ψηφιοποίηση και περιλαμβάνει τα στάδια της δειγματοληψίας και της κβάντισης.

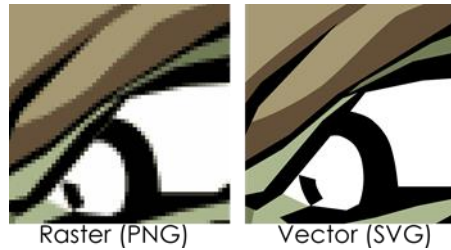
Έχοντας κατανοήσει τις παραπάνω βασικές έννοιες θα ήταν χρήσιμο εάν δίναμε τον ορισμό της ψηφιακής εικόνας. Μια ψηφιακή εικόνα είναι μια αριθμητική αναπαράσταση (συνήθως δυαδική) από μια δισδιάστατη εικόνα. Ανάλογα με την σταθερότητα της ανάλυσης της εικόνας η ψηφιακή εικόνα κατηγοριοποιείται σε vector και raster εικόνες. Συνήθως οι ψηφιακές εικόνες αναφέρονται σε raster εικόνες τις οποίες τις συναντάμε και ως εικόνες bitmap.

1.2 Raster και Vector εικόνες

Οι raster εικόνες έχουν ένα πεπερασμένο σύνολο ψηφιακών τιμών, τα γνωστά pixels. Η ψηφιακή εικόνα περιέχει ένα σταθερό αριθμό γραμμών και στηλών των pixels. Τυπικά, τα pixels που αποθηκεύονται στη μνήμη του υπολογιστή αποθηκεύονται ως μια raster εικόνα, μια δισδιάστατη διάταξη των μικρών ακεραίων. Οι τιμές αυτές συχνά μεταδίδονται ή αποθηκεύονται σε συμπιεσμένη μορφή.

Οι raster εικόνες μπορούν να δημιουργηθούν από μια ποικιλία συσκευών εισόδου και τεχνικές, όπως ψηφιακές φωτογραφικές μηχανές, σαρωτές, μηχανές μέτρησης, σειсмоγράφους, ραντάρ, και πολλά άλλα. Μπορούν επίσης να συντεθούν από αυθαίρετα μη εικονικά δεδομένα, όπως μαθηματικές συναρτήσεις ή τρισδιάστατα γεωμετρικά μοντέλα. Το τελευταίο είναι μια σημαντική υποπεριοχή των γραφικών υπολογιστών. Το πεδίο της ψηφιακής επεξεργασίας εικόνας είναι η μελέτη των αλγορίθμων για τον μετασχηματισμό τους.

Από την άλλη οι vector εικόνες προέκυψαν από τη μαθηματική γεωμετρία (vector). Σε μαθηματικούς όρους, ένα διάνυσμα αποτελείται από το σημείο το οποίο έχει τόσο την κατεύθυνση όσο και το μήκος.



Εικόνα 1.1 Αριστερά:εικόνα raster Δεξιά:εικόνα vector

Συχνά, τόσο τα raster και τα vector στοιχεία συνδυάζονται σε μία εικόνα. Για παράδειγμα, στην περίπτωση μιας πινακίδας με κείμενο (vector) και στην περίπτωση φωτογραφίας (raster).

Η χρησιμότητα της ψηφιακής εικόνας φαίνεται και μέσα από τις διάφορες εφαρμογές της όπως κάποιες από αυτές θα περιγραφούν παρακάτω.

1.3 Διάφορες Εφαρμογές και πρότυπα Ψηφιακής εικόνας

Κάνοντας μια γρήγορη ιστορική αναδρομή οι πρώτες ψηφιακές συσκευές fax όπως το Bartlane προηγήθηκαν πολλές δεκαετίες πριν τις ψηφιακές φωτογραφικές μηχανές και τους υπολογιστές. Η πρώτη εικόνα που σαρώθηκε, αποθηκεύτηκε, και αναδημιουργήθηκε σε ψηφιακά pixels είχε εμφανιστεί στα πρότυπα της Ανατολικής Automatic Computer (SEAC) στο NIST. Η προώθηση της ψηφιακής απεικόνισης συνεχίστηκε στις αρχές του 1960, παράλληλα με την ανάπτυξη του διαστημικού προγράμματος και της ιατρικής έρευνας. Εργασίες στο Jet Propulsion Laboratory, MIT, Bell Labs και το Πανεπιστήμιο του Maryland, μεταξύ άλλων, χρησιμοποίησαν ψηφιακές εικόνες για την προώθηση δορυφορικών εικόνων, τηλεγραφική μετατροπή φωτογραφίας, ιατρική απεικόνιση, την τεχνολογία εικονοτηλεφώνου, αναγνώρισης χαρακτήρων, και τη βελτίωση της φωτογραφίας.

Η ταχεία πρόοδος στην ψηφιακή απεικόνιση ξεκίνησε με την εισαγωγή των μικροεπεξεργαστών στις αρχές του 1970. Η εφεύρεση της αξονικής τομογραφίας (CAT scanning), χρησιμοποίησε ακτίνες X για να παράγει μία ψηφιακή εικόνα ενός "κομματιού" μέσα από ένα τρισδιάστατο αντικείμενο, είχε μεγάλη σημασία για τις ιατρικές διαγνώσεις. Όπως επίσης και την προέλευση των ψηφιακών εικόνων, την ψηφιοποίηση αναλογικών εικόνων όπου επιτρέπεται η βελτίωση και η αποκατάσταση των αρχαιολογικών αντικειμένων.

Οι πρόοδοι στην τεχνολογία των μικροεπεξεργαστών άνοιξαν το δρόμο για την ανάπτυξη και την εμπορία των CCDs συσκευών για χρήση σε ένα ευρύ φάσμα των συσκευών καταγραφής εικόνας. Σταδιακά εξαφανίστηκε η χρήση των αναλογικών ταινιών και της βιντεοσκόπησης με ταινίες στα τέλη του 20^{ου} αιώνα. Τέλος αξίζει να σημειωθεί πως απαιτείται μεγάλη υπολογιστική ισχύς για την επεξεργασία της ψηφιακής εικόνας.

Η συμπίεση εικόνας είναι απαραίτητη εκτός από την ανάγκη μετάδοσης της εικόνας και στην περίπτωση που χρειάζεται μία προεπισκόπηση εικόνων που ανήκουν σε μία μεγάλη βάση δεδομένων [2]. Η αρχή της συμπίεσης της εικόνας είναι να επιλεγούν οι πληροφορίες που θέλουμε να κρατήσουμε και να τις κωδικοποιήσουμε χωρίς απώλειες.

Επομένως το πρώτο στάδιο κατά τη διαδικασία της συμπίεσης είναι να βρεθεί η ελάχιστη ομάδα από κβαντισμένους συντελεστές που θα δώσουν μία εικόνα με υψηλή ποιότητα. Οι πληροφορίες που χάνονται δε μπορούν να ανακτηθούν για αυτό το λόγο αν δεν επιθυμούμε να έχουμε καθόλου απώλειες θα συμπίεσουμε και τον θόρυβο.

Κατά την ανάκτηση της αρχικής εικόνας με τη χρήση της συμπιεσμένης είναι δυνατόν να μη χρειαστεί ολόκληρη η συμπιεσμένη εικόνα. Αυτό συμβαίνει γιατί η αποσυμπίεση γίνεται κατά κλίμακες ξεκινώντας από χαμηλή ανάλυση και αυξάνοντας τη σιγά σιγά ανάλογα με την επιθυμία του παραλήπτη. Αν ο παραλήπτης επιθυμεί απλά να δει το θέμα της εικόνας δεν είναι απαραίτητο να αποσυμπίεσει την εικόνα στην αρχική της ανάλυση.

Η ανίχνευση των αντικειμένων είναι μία τεχνολογία των ηλεκτρονικών υπολογιστών που σχετίζεται με την τεχνητή όραση και την ψηφιακή επεξεργασία εικόνας. Αυτή η τεχνολογία ασχολείται με την ανίχνευση αντικειμένων σε ψηφιακές εικόνες και βίντεο. Χαρακτηριστικά παραδείγματα τέτοιων εφαρμογών είναι συστήματα ανίχνευσης προσώπων και ανθρώπων γενικότερα. Τέτοιου είδους συστήματα βρίσκουν εφαρμογή σε διάφορους τομείς όπως η παρακολούθηση μέσω βίντεο.



Εικόνα 1.2 Παράδειγμα ανίχνευσης αντικειμένων

Οι διαφορετικοί τρόποι απεικόνισης της ψηφιακής εικόνας εξυπηρετούν σε διάφορες εφαρμογές καθώς κάνουν πιο εύκολη την εξαγωγή πληροφοριών μιας εικόνας ανάλογα με την εφαρμογή. Υπάρχει η έννοια των σημασιολογικών χαρακτηριστικών της εικόνας τα οποία μπορούν να βρεθούν όταν η απεικόνιση

της εικόνας έχει μια συγκεκριμένη δομή και μπορούν να δώσουν μία ερμηνεία της εικόνας.

Σε εικόνες οι οποίες έχουν μεγάλα αντικείμενα χρησιμοποιήθηκαν συγκεκριμένες τεχνικές ώστε να ανιχνευτούν από διάφορους επιστήμονες όπως (3), (4), (5). Υπάρχουν όμως και οι περιπτώσεις όπου είναι απαραίτητη η ανίχνευση πολύ μικρών αντικειμένων σε μία εικόνα όπως αστρονομικά αντικείμενα.

Για τον εντοπισμό και την ανάλυση συγκεκριμένων αντικειμένων χρησιμοποιούνται μετασχηματισμοί πολλαπλής ανάλυσης της εικόνας. Υπάρχουν χάρτες των αντικειμένων της εικόνας ή χρησιμοποιείται μία δυαδική απεικόνιση της εικόνας η οποία μπορεί να επεξεργαστεί περαιτέρω ώστε να απορρίφθουν τυχόν λανθασμένες ανιχνεύσεις αντικειμένων. Κάποιες από τις τεχνικές που χρησιμοποιούνται σε τέτοιες περιπτώσεις βασίζονται σε διαχειρισή τους με βάση τη μαθηματική μορφολογία Minkowski (6).

Έχουν γίνει πολλές μελέτες για να αποδείξουν ότι το περιεχόμενο της εικόνας παίζει βασικό ρόλο στην ανίχνευση των αντικειμένων που αυτή απεικονίζει (7), (8). Το περιεχόμενο της εικόνας είναι ένας όρος που χρησιμοποιείται συχνά στην τεχνητή όραση χωρίς να υπάρχει ένας σαφής ορισμός. Ουσιαστικά αναφέρεται σε κάθε πληροφορία της εικόνας που μπορεί να επηρεάσει τη σκηνή και τα αντικείμενα που απεικονίζονται.

Μία έξυπνη εφαρμογή – εκμετάλευση της τοπικής πληροφορίας, δηλαδή του περιεχομένου συγκεκριμένων pixel χρησιμοποιείται στην ανίχνευση αντικειμένων όπου αρχικά τοποθετείται ένα παράθυρο σε μία μικρή περιοχή του αντικειμένου και στη συνέχεια αυτό μεγαλώνει σιγά σιγά περιλαμβάνοντας γειτονικά pixel τα οποία έχουν παρόμοιο περιεχόμενο (9), (10).

Η τοπική πληροφορία του περιεχομένου των pixel είναι μία τεχνική που ρησιμοποιείται στην κατάτμηση εικόνων, στην εξαγωγή των ορίων των αντικειμένων και σε μοντέλα αντικειμένων/περιγραμμάτων. Το περιεχόμενο μιας εικόνας χωρίζεται σε κάποιες κατηγορίες. Κάποιες από αυτές περιγράφονται παρακάτω και περιλαμβάνουν το τριδιάστατο περιεχόμενο, το σημασιολογικό, το φωτογραφικό, το περιεχόμενο του φωτός, το καιρικό, το γεωγραφικό, το χρονικό και το πολιτισμικό (11).

Το γεωμετρικό περιεχόμενο μία εικόνας βοηθάει στην τριδιάστατη αναπαράσταση της σκηνης που απεικονίζεται ώστε να αποκαλυφθεί η διάταξη των επιφανειών και των διαφορετικών αντικειμένων δίνοντας απαντήσεις σε προβλήματα που μπορεί να προκληθούν για παράδειγμα από την αλληλοεπικάλυψη των αντικειμένων (12).

Η δημιουργία ενός ψηφιακού κειμένου, μιας ψηφιακής εικόνας, ενός ψηφιακού ήχου ή ενός ψηφιακού βίντεο περικλείεται στον όρο ψηφιοποίηση έργου. Η ψηφιοποίηση περιλαμβάνει μία σειρά από στάδια από το αρχικό μέχρι το τελικό αποτέλεσμα: σχεδιασμός του έργου ψηφιοποίησης, επιλογή περιεχομένου, προετοιμασία για ψηφιοποίηση, μεταχείριση των πρωτοτύπων, ψηφιοποίηση, διατήρηση του ψηφιακού περιεχομένου, μεταδεδομένα, ενέργειες ανάδειξης – προβολής, πνευματικά δικαιώματα και διαχείριση έργων ψηφιοποίησης[13][14].

Το πιο βασικό στάδιο είναι αυτό της ψηφιοποίησης που περιλαμβάνει τη σάρωση, την ψηφιακή φωτογράφιση και την ψηφιακή αποτύπωση των πρωτοτύπων συμπεριλαμβάνοντας και την οποιαδήποτε ψηφιακή επεξεργασία που μπορεί να έχει γίνει.

Τα πρότυπα της εικόνας και των μεταδεδομένων είναι μία κρίσιμη απόφαση και δεν πρέπει να προσπερνιέται. Τα τελευταία χρόνια έχουν επικρατήσει συγκεκριμένα πρότυπα τα οποία περιγράφονται με συντομία στη συνέχεια. Όταν ο χρήστης έχει στην κατοχή του μία εικόνα ενός συγκεκριμένου προτύπου μπορεί να τη μετατρέψει σε άλλο πρότυπο με το κατάλληλο λογισμικό επεξεργασίας εικόνας.

Έχοντας κατανοήσει τι είναι ψηφιακή εικόνα και ποιες οι εφαρμογές της θα πρέπει να αναφερθούν ακόμη και τα 4 κύρια πρότυπα (formats) των ψηφιακών εικόνων.

TIFF (γνωστές και ως TIF), η κατάληξη των αρχείων αυτών είναι .tif

Οι εικόνες τύπου TIFF είναι συντομογραφία της Tagged Image File Format. Οι εικόνες αυτές δημιουργούν πολύ μεγάλα μεγέθη αρχείων όπως επίσης είναι ασυμπίεστες και έτσι περιέχουν πολλά λεπτομερή δεδομένα. Οι TIFF είναι εξαιρετικά ευέλικτες όσον αφορά το χρώμα και το περιεχόμενο.

Επιπροσθέτως είναι ο πιο κοινός τύπος αρχείου που χρησιμοποιείται από το φωτογραφικό λογισμικό (Photoshop).

JPEG (γνωστά και ως JPG), η κατάληξη των αρχείων αυτών είναι .jpg

Το JPEG σημαίνει Joint Photographic Experts Group και χρησιμοποιείται ευρέως. Χρησιμοποιείται για τη μεταφορά και την παρουσίαση εικόνων μέσω δικτύων τα οποία έχουν μικρό εύρος ζώνης δεδομένου ότι οι εικόνες προτύπου JPEG καταλαμβάνουν μικρό όγκο. Σε αντίθεση με το πρότυπο TIFF κατά τη συμπίεση με βάση το JPEG υπάρχει απώλεια όγκου η οποία μειώνει τη χωρητικότητα της εκάστοτε εικόνας.

GIF, με κατάληξη .gif

Η GIF είναι η συντομογραφία του Graphic Interchange Format. Η συμπίεση κατά το πρότυπο GIF γίνεται χωρίς απώλεια πληροφορίας αλλά μειώνει τον όγκο της εικόνας. Ένα μειονέκτημα αυτού του προτύπου είναι ότι υποστηρίζει μόλις 256 χρώματα. Το GIF χρησιμοποιείται σε γραφικά. Οι φυλλομετρητές ιστού και η πλειοψηφία των εφαρμογών ηλεκτρονικού υπολογιστή υποστηρίζουν το συγκεκριμένο πρότυπο.

PNG, με κατάληξη .png

Portable Network Graphics ή αλλιώς PNG είναι ένα ανοικτό πρότυπο και ο στόχος ανάπτυξής του ήταν να αντικαταστήσει το πρότυπο GIF. Κατά τη συμπίεση ο όγκος μειώνεται αλλά η πληροφορία δε χάνεται και συγκριτικά με το GIF το μέγεθος των εικόνων είναι μικρότερο.

Στην παρούσα διπλωματική το κύριο συστατικό είναι η ψηφιακή εικόνα, αλλά πως μπορούμε να επεξεργαστούμε μία τέτοια εικόνα χωρίς το απαραίτητο λογισμικό.

Παρακάτω περιγράφεται εν συντομία το υπολογιστικό περιβάλλον στο οποίο έγινε η οποιαδήποτε επεξεργασία

1.4 Τι είναι το Matlab

Το MATLAB (MATrixLABoratory) είναι ένα αριθμητικό υπολογιστικό περιβάλλον λογισμικού στους ηλεκτρονικούς υπολογιστές το οποίο ανήκει στις γλώσσες προγραμματισμού τέταρτης γενιάς. Χρησιμοποιείται ευρέως από επιστήμονες που αναπτύσσουν συστήματα λογισμικών βασισμένα στα μαθηματικά.

Αναπτύχθηκε από την MathWorks και πρόκειται για ένα πρόγραμμα λογισμικού που η ανάπτυξή του ξεκίνησε ως ένα εργαστήριο πινάκων παρέχοντας μεγάλη διευκόλυνση σε υπολογισμούς μεταξύ πινάκων αλλά και στην αποθήκευση μεγάλων δεδομένων σε πίνακες κάνοντας εύκολη της οργάνωσή τους, την προσπέλαση τους και την επεξεργασία τους μαζικά.

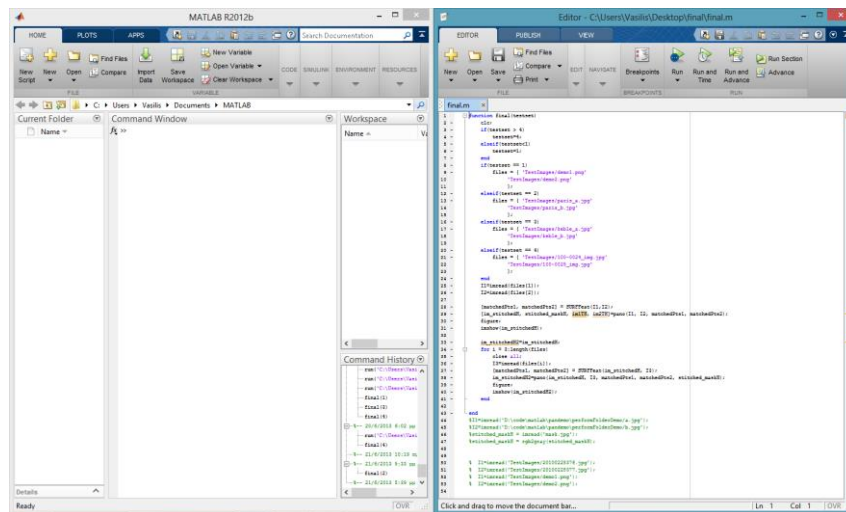
Θεωρείται ένα από τα πιο ισχυρά εργαλεία σε πολλούς τομείς όπως αυτοί της οπτικοποίησης, του προγραμματισμού, της έρευνας, των μηχανικών συστημάτων και των επικοινωνιών. Στα πλαίσια του MATLAB έχουν αναπτυχθεί μοντέρνοι αλγόριθμοι και ισχυρά προγραμματιστικά εργαλεία. Το MATLAB αποτελείται από το βασικό πακέτο συναρτήσεων, βιβλιοθηκών και εργαλείων καθώς και από ανεξάρτητες βιβλιοθήκες και εργαλειοθήκες που έχουν αναπτυχθεί και αναπτύσσονται και παρέχονται ξεχωριστά.

Κάποιος χρήστης ανάλογα με το σύστημα που θέλει να δημιουργήσει χρησιμοποιώντας το MATLAB θα μπορέσει να διαχειριστεί πίνακες, να κάνει διαγράμματα συναρτήσεων και δεδομένων, να υλοποιήσει αλγορίθμους, να δημιουργήσει διεπαφές χρήστη καθώς και να συνδέσει προγράμματα ανεπτυγμένα στο MATLAB με προγράμματα άλλων γλωσσών προγραμματισμού (C, C++, Java και Fortran).

Ιστορία: Ο Cleve Moler, πρόεδρος του τμήματος επιστήμης των υπολογιστών στο Πανεπιστήμιο του Νέου Μεξικού, ξεκίνησε την ανάπτυξη του MATLAB στα τέλη της δεκαετίας του 1970. Ουσιαστικά το σχεδίασε έτσι ώστε να δώσει στους

μαθητές του πρόσβαση στις Linpack και Eispack (βιβλιοθήκες) χωρίς να χρειάζεται να γνωρίζουν την FORTRAN. Γρήγορα εξαπλώθηκε σε άλλα πανεπιστήμια και βρήκε ένα ισχυρό κοινό μέσα στην κοινότητα των εφαρμοσμένων μαθηματικών. Ένας από αυτούς ήταν ο μηχανικός Jack Little όπου αργότερα μαζί με τον Moler και τον Steve Bangert ανέπτυξαν την MATLAB από την αρχή με τη γλώσσα προγραμματισμού C και ίδρυσαν την Mathworks το 1984 για να συνεχιστεί μετέπειτα η ανάπτυξη της. Το 2000, το MATLAB αναπτύχθηκε ξανά χρησιμοποιώντας πλέον μια νεότερη σειρά βιβλιοθηκών την LAPACK.

Από τότε άρχισαν να το χρησιμοποιούν επιστήμονες των μηχανικών επιστημών αλλά πολλοί γρήγορα διαδόθηκε και άρχισε να χρησιμοποιείται και σε άλλους τομείς. Τέλος είναι ευρέως διαδεδομένο για την επεξεργασία εικόνας.



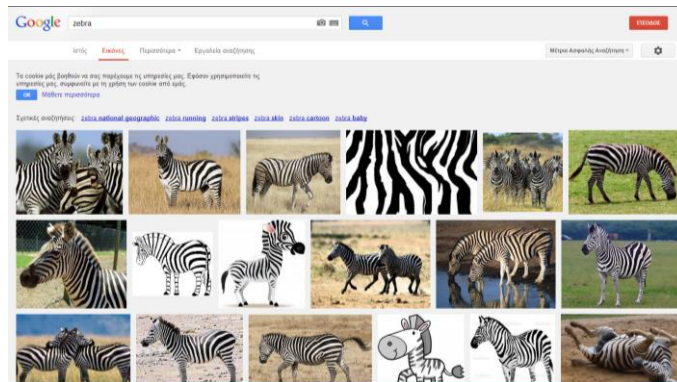
Εικόνα 1.3 Γραφικό περιβάλλον της εφαρμογής MATLAB

Κεφάλαιο 2.

Αλγόριθμοι SURF & SIFT

2.1 Εισαγωγή

Καθημερινά εκατομμύρια άνθρωποι έρχονται σε επαφή με τον προσωπικό τους υπολογιστή και κατά συνέπεια με το διαδίκτυο, όπου συναντούν ένα τεράστιο αριθμό εικόνων. Πιο συγκεκριμένα μέσω της μεγαλύτερης μηχανής αναζήτησης ιστού το Google μας δίνεται η δυνατότητα να ανακτήσουμε εικόνες από το ίδιο σημασιολογικό περιβάλλον. Π.χ πληκτρολογώντας τη λέξη zebra στην μηχανή αναζήτησης μας εμφανίζεται η παρακάτω εικόνα :



Εικόνα 2.1 Αποτέλεσμα εικόνων ύστερα απο αναζήτηση στο Google

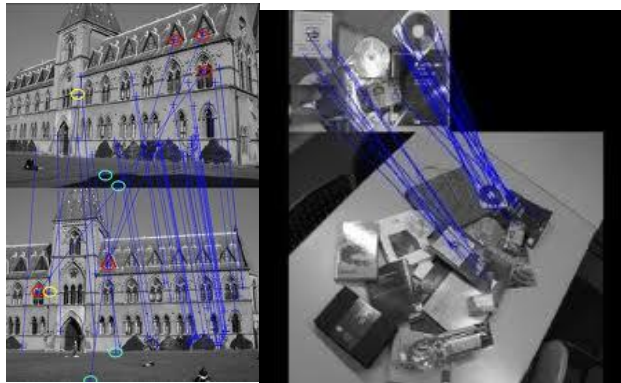
Όλες αυτές οι εικόνες είναι, κοινού σημασιολογικού περιβάλλοντος, το ερώτημα μας είναι μπορεί να γίνει αντιστοίχιση των εικόνων (image matching) με κάποια τεχνική; Η αντιστοίχιση εικόνων έχει σαν σκοπό την εύρεση γεωμετρικών και άλλων διαφορών ανάμεσα σε δύο ή περισσότερες εικόνες. Η Ανίχνευση, ο εντοπισμός και η αναγνώριση εξέχοντων περιοχών ή χαρακτηριστικών σημείων σε εικόνες είναι ένα πολύ σημαντικό και θεμελιώδες πρόβλημα στην τεχνητή όραση υπολογιστών αλλά και της ρομποτικής κοινότητας.

Για να γίνει η αντιστοίχιση δύο ή περισσότερων εικόνων απαιτείται να γίνει πρώτα αντιστοίχιση χαρακτηριστικών (feature matching), η οποία είναι η βάση πολλών προβλημάτων της τεχνητής όρασης, όπως η αναγνώριση αντικειμένου ή της δομής. Οι τρέχουσες μέθοδοι βασίζονται σε περιγραφείς για την ανίχνευση και την αντιστοίχιση σημείων.

Η απάντηση στο παραπάνω ερώτημα δίνεται παρακατω. Οι δύο πιο πρόσφατες αλλά και υποσχόμενες προσεγγίσεις-τεχνικές που έχουν δημοσιευθεί για την

ανίχνευση κύριων περιοχών στις εικόνες είναι οι : **SIFT**, Lowe (2004) (Scale Invariant Feature Transform) και **SURF**, Bay (2006) (Speeded Up Robust Features). Και οι δύο προσεγγίσεις δεν εντοπίζουν μόνο σημεία ενδιαφέροντος(interest points-τα λεγόμενα χαρακτηριστικά της εικόνας) αλλά προτείνουν και μια μέθοδο για την δημιουργία ενός αμετάβλητου περιγραφέα. Αυτός ο περιγραφέας μπορεί να χρησιμοποιηθεί (περισσότερο ή λιγότερο) προσδιορίζοντας με μοναδικό τρόπο τα σημεία ενδιαφέροντος τα οποία έχουν βρεθεί και αντιστοιχίζοντας τα ακόμη και κάτω από μια ποικιλία διαταραχόμενων συνθηκών , όπως αλλαγές κλίμακας, περιστροφή, αλλαγές στο φωτισμό ή στο θόρυβο της εικόνας.

Οι δύο αυτοί αλγόριθμοι μοιράζονται κοινούς σκοπούς στην αναγνώριση προτύπων και χρησιμοποιούνται ευρέως στην πρόσφατη βιβλιογραφία. Για την εμφάνιση, τα τελευταία χρόνια χρησιμοποιείται ευρέως ο αλγόριθμος SIFT (Scale Invariant Feature Transform) και λιγότερο άλλοι αλγόριθμοι όπως οι SURF (Speeded Up Robust Features), SPIN και RIFT. Οι παραπάνω αλγόριθμοι εντοπίζουν στις εικόνες σημεία-κλειδιά (key points) και στη συνέχεια περιγράφονται με διανύσματα τα χαρακτηριστικά της εικόνας σε περιοχές με κέντρα τα σημεία-κλειδιά .



Εικόνα 2.2 Αριστερά: Παράδειγμα αντιστοίχισης σημείων με SIFT, Δεξιά: Παράδειγμα αντιστοίχισης σημείων με SURF

2.2 Αλγόριθμος SIFT (Scale-invariant feature transform)

Όπως αναφέρθηκε παραπάνω ο αλγόριθμος SIFT είναι πιο διαδεδομένος και χρησιμοποιείται πιο σε σχέση με τον αλγόριθμο SURF. Ο SIFT πάνω από μια δεκαετία έχει αποδειχθεί εξαιρετικά επιτυχημένος σε μια σειρά εφαρμογών που χρησιμοποιούν οπτικά χαρακτηριστικά, όπως η αναγνώριση του αντικειμένου (object recognition), 'συρραφή εικόνας' (image stitching), οπτική χαρτογράφηση (visual mapping) κλπ. Στις επόμενες ενότητες θα αναλύσουμε λεπτομερώς τον αλγόριθμο SIFT.

Ο SIFT είναι η συντομογραφία του 'Scale-invariant feature transform' και είναι ο αλγόριθμος ο οποίος δημιουργήθηκε έτσι ώστε να εντοπίζει και να περιγράφει τοπικά χαρακτηριστικά σε εικόνες. Ο αλγόριθμος δόθηκε στη δημοσιότητα για πρώτη φορά από τον David Lowe το 1999 και είναι κατοχυρώμενος με στις ΗΠΑ και πιο συγκεκριμένα στο University of British Columbia.

Οι εφαρμογές του περιλαμβάνουν αναγνώριση αντικειμένων, ρομποτική χαρτογράφηση και πλοήγηση, συρραφή εικόνας, 3D μοντελοποίηση, αναγνώριση χειρονομιών και παρακολούθηση βίντεο.

2.2.1 Γενική επισκόπηση του αλγορίθμου

Για κάθε αντικείμενο σε μία εικόνα, τα σημεία ενδιαφέροντος του αντικειμένου μπορούν να εξαχθούν για να 'παράγουν' μια "περιγραφή χαρακτηριστικών" του αντικειμένου. Αυτή η περιγραφή, προέρχεται από μια εικόνα κατάρτισης (training image), στη συνέχεια μπορεί να χρησιμοποιηθεί για την αναγνώριση του αντικειμένου κατά την προσπάθεια της να εντοπίσει το αντικείμενο σε μια δοκιμασική εικόνα η οποία περιέχει πολλά άλλα αντικείμενα. Για την εκτέλεση μίας αξιόπιστης αναγνώρισης, είναι σημαντικό τα χαρακτηριστικά τα οποία εξάγονται από την εικόνα κατάρτισης να είναι ανιχνεύσιμα ακόμη και κάτω από αλλαγές στην κλίμακα της εικόνας, του θορύβου και του φωτισμού. Αυτά τα σημεία συνήθως βρίσκονται σε περιοχές υψηλής αντίθεσης της εικόνας, όπως τα άκρα των αντικειμένων

Ένα άλλο σημαντικό γνώρισμα αυτών των χαρακτηριστικών είναι ότι οι σχετικές θέσεις μεταξύ τους στην αρχική σκηνή δεν θα πρέπει να αλλάζουν από τη μία εικόνα στην άλλη. Για παράδειγμα, εάν μόνον οι τέσσερις γωνίες μιας πόρτας χρησιμοποιήθηκαν ως χαρακτηριστικά, θα μπορούσαν να λειτουργήσουν ανεξάρτητα από τη θέση της πόρτας. Αλλά εάν και τα σημεία στο πλαίσιο χρησιμοποιήθηκαν, η αναγνώριση θα αποτύχει εάν η πόρτα είναι ανοιχτή ή κλειστή. Ομοίως, τα χαρακτηριστικά τα οποία βρίσκονται σε αρθρωτά ή εύκαμπτα αντικείμενα δεν θα λειτουργούν εάν συμβεί οποιαδήποτε αλλαγή στην

εσωτερική τους γεωμετρία μεταξύ δύο εικόνων στο σύνολο, υπό επεξεργασία. Ωστόσο, στην πράξη ο SIFT ανιχνεύει και χρησιμοποιεί ένα πολύ μεγαλύτερο αριθμό χαρακτηριστικών από τις εικόνες, ο οποίος μειώνει τη συμβολή των σφαλμάτων που προκαλούνται από τις τοπικές διακυμάνσεις στο μέσο σφάλμα σε σχέση με όλα τα χαρακτηριστικά σφάλματα αντιστοίχισης.

2.2.2 Βασικά στάδια αλγόριθμου

Σε αυτήν την ενότητα θα αναλυθούν τα 5 βασικά στάδια του αλγόριθμου SIFT[15].

- **Αμετάβλητη-κλίμακα ανίχνευσης χαρακτηριστικών(Scale-invariant feature detection)**

Η μέθοδο του Lowe για την παραγωγή χαρακτηριστικών μίας εικόνας μετατρέπει μια εικόνα σε μια μεγάλη συλλογή χαρακτηριστικών διανυσμάτων, καθένα από τα οποία είναι αμετάβλητο σε μετάφραση εικόνας, κλιμάκωση και περιστροφή, μερικώς αμετάβλητο σε αλλαγές φωτισμού και ισχυρο στην τοπική γεωμετρική παραμόρφωση. Οι βασικές περιοχές ορίζονται ως μέγιστα και ελάχιστα για το αποτέλεσμα της διαφοράς της Γκαουσιανής συνάρτησης(Gaussian function) που ισχύει στον κλιμακωτό χώρο σε μια σειρά από ομαλές εικόνες και εικόνες αναδειγματοληψίας. Ακόμη απορρίπτονται τα χαμηλής αντίθεσης σημεία και τα σημεία ανταπόκρισης κατά μήκος μιας ακμής. Οι κυρίαρχες κατευθύνσεις ανατείνονται στα τοπικά σημεία-κλειδιά(keypoints). Αυτά τα βήματα εξασφαλίζουν ότι τα σημεία-κλειδιά είναι πιο σταθερά για το ταίριασμα και την αναγνώριση.

- **Αντιστοίχιση χαρακτηριστικών και αποδελτίωση(Feature matching and indexing)**

Η αποδελτίωση αποτελείται από την αποθήκευση SIFT κλειδιών και τον εντοπισμό κλειδιών αντιστοίχισης από τη νέα εικόνα. Ο Lowe χρησιμοποίησε μια τροποποίηση του αλγορίθμου k-d δέντρο (k-d tree) που ονομάζεται Best-bin-first search(BBF) η οποία μπορεί να εντοπίσει τους πλησιέστερους γείτονες με υψηλή πιθανότητα, χρησιμοποιώντας μόνο ένα περιορισμένο αριθμό των υπολογισμών. Ο αλγόριθμος BBF χρησιμοποιεί μια τροποποιημένη αναζήτηση του αλγορίθμου k-d έτσι ώστε οι κάδοι των χαρακτηριστικών χώρου να έχουν αναζητηθεί με σειρά πλησιέστερης απόστασης τους από την θέση ερωτήματος(query location). Αυτή η σειρά αναζήτησης απαιτεί τη χρήση μίας ουράς προτεραιότητας(heap-based priority) για τον αποτελεσματικό

προσδιορισμό της αναζήτησης. Ο καλύτερη υποψήφια αντιστοίχιση για κάθε σημείο-κλειδί βρίσκεται με τον εντοπισμό του πλησιέστερου γείτονα στη βάση δεδομένων των σημείων-κλειδιών των εικόνων κατάρτισης (training images). Οι πλησιέστεροι γείτονες ορίζονται ως τα σημεία-κλειδιά με την ελάχιστη Ευκλείδεια απόσταση από το δοσμένο διάνυσμα του Περιγραφέα. Η πιθανότητα μία αντιστοίχιση να είναι σωστή μπορεί να προσδιοριστεί λαμβάνοντας την αναλογία της απόστασης από τον πλησιέστερο γείτονα με την απόσταση του δεύτερου πλησιέστερου.

Ο Lowe απέρριψε όλες τις αντιστοιχίες στις οποίες ο λόγος της απόστασης είναι μεγαλύτερος από 0,8, ο οποίος εξαλείφει το 90% των εσφαλμένων αντιστοιχίσεων, ενώ απορρίπτει λιγότερο από το 5% των σωστών αντιστοιχίσεων. Για την περαιτέρω βελτίωση της αποτελεσματικότητας του BBF, η αναζήτηση αποκόπτεται μετά τον έλεγχο των πρώτων 200 υποψηφίων πλησιέστερο γειτόνων.

- **Ταυτοποίηση συστάδας-συνόλου από τον μετασχηματισμό Hough(Cluster identification by Hough transform voting)**

Ο μετασχηματισμός Hough (Hough Transform) χρησιμοποιείται για να ομαδοποιεί αξιόπιστα μοντέλα υποθέσης έτσι ώστε να αναζητούνται κλειδιά τα οποία συμφωνούν με ένα συγκεκριμένο μοντέλο . Ο μετασχηματισμός αυτός προσδιορίζει σύνολα χαρακτηριστικών με μια συνεπή ερμηνεία. Όταν τα σύνολα των χαρακτηριστικών που βρέθηκαν για να επιλέξουν για την ίδια στάση ενός αντικειμένου, η πιθανότητα της ερμηνείας να είναι σωστή είναι πολύ υψηλότερη από ό, τι για κάθε μεμονωμένο χαρακτηριστικό. Μια καταχώρηση σε έναν πίνακα κατακερματισμού δημιουργείται από την πρόβλεψη της θέσης του μοντέλου , τον προσανατολισμό και την κλίμακα υπόθεσης αντιστοίχισης .Ο πίνακας κατακερματισμού είναι η αναζήτηση για τον εντοπισμό όλων των ομάδων τουλάχιστον 3 καταχωρήσεων σε έναν κάδο, και οι κάδοι ταξινομούνται σε φθίνουσα σειρά ανάλογα το μέγεθός τους.

- **Επαλήθευση μοντέλου με γραμμική λύση ελαχίστων τετραγώνων (Model verification by linear least squares)**

Κάθε εντοπισμένο σύνολο υπόκειται στη συνέχεια σε μια διαδικασία εξακρίβωσης κατά την οποία μια γραμμική λύση ελαχίστων τετραγώνων εκτελείται για τις παράμετρους του μετασχηματισμού affine που αφορούν το μοντέλο στην εικόνα. Ο μετασχηματισμός affine ενός σημείου μοντέλου $[x \ y]^T$ σε ένα σημείο της εικόνας $[u \ v]^T$ μπορεί να γραφεί ως εξής:

$$\begin{matrix} u \\ v \end{matrix} = \begin{matrix} m1 & m2 \\ m3 & m4 \end{matrix} \cdot \begin{matrix} x \\ y \end{matrix} + \begin{matrix} tx \\ ty \end{matrix} \quad (2.1)$$

Όπου η μετάφραση του μοντέλου είναι η $[tx \ ty]^T$ και η affine περιστροφή, κλίμακα και έκταση παρουσιάζονται από τις παραμέτρους $m1, m2, m3$, και $m4$. Για την επίλυση των παραμέτρων του μετασχηματισμού η παραπάνω εξίσωση μπορεί να ξαναγραφεί για να συγκεντρώσει τους αγνώστους σε ένα διάνυσμα-στήλη.

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ \dots & & & & & \\ \dots & & & & & \end{bmatrix} \begin{bmatrix} m1 \\ m2 \\ m3 \\ m4 \\ tx \\ ty \end{bmatrix} = \begin{bmatrix} u \\ v \\ . \\ . \end{bmatrix} \quad (2.2)$$

Η παραπάνω εξίσωση δείχνει μάλιστα αντιστοίχιση, αλλά μπορεί να προστεθεί ο οποιοσδήποτε αριθμός περαιτέρων αντιστοιχίσεων. Με κάθε αντιστοίχιση δύο ακόμη γραμμές προστίθενται στον πρώτο και τον τελευταίο πίνακα. Ο μικρότερος δυνατός αριθμός που απαιτείται για να δοθεί μία λύση είναι 3. Το παραπάνω γραμμικό σύστημα μπορούμε να το γράψουμε ως εξής:

$$A\hat{x} \approx b \quad (2.3)$$

Όπου A είναι ο m -by- n πίνακας (συνήθως $m > n$), x ένα άγνωστο n -διάστατο διάνυσμα παραμέτρων, και το b ένα m -διάστατο διάνυσμα μέτρησης. Συνεπώς, το ελαχιστοποιημένο διάνυσμα είναι μία λύση της κανονικής εξίσωσης.

$$A^T A \hat{x} = A^T b \quad (2.4)$$

- **Ανίχνευση των outliers (Outlier detection)**

Τα outliers μπορούν πλέον να αφαιρεθούν ελέγχοντας εάν υπάρχει αντιστοίχιση-συμφωνία μεταξύ κάθε χαρακτηριστικού της εικόνας και του μοντέλου, δεδομένης λύσης παραμέτρου. Λαμβάνοντας υπόψη τη

γραμμική λύση ελαχίστων τετραγώνων, κάθε αντιστοίχιση οφείλει να συμφωνεί με το μισό εύρος σφάλματος που χρησιμοποιήθηκε για τις παραμέτρους των κάδων του μετασχηματισμού Hough. Ως outliers απορρίπτονται, η γραμμική λύση ελαχίστων τετραγώνων δίνει εκ νέου λύση με τα υπόλοιπα σημεία, και η διαδικασία επαναλαμβάνεται. Αν έχουν απομείνει λιγότερα από 3 σημεία τότε παραμένουν και μετά την απόρριψη των outliers και τότε η αντιστοίχιση απορρίπτεται.

Η τελική απόφαση για την αποδοχή ή απόρριψη μιας υπόθεσης μοντέλου βασίζεται σε ένα λεπτομερές πιθανοτικό μοντέλο (Wikipedia). Αυτή η μέθοδος υπολογίζει πρώτα το αναμενόμενο αριθμό των ψευδών αντιστοιχίσεων στο μοντέλο, δεδομένου του προβλεπόμενου μεγέθους του μοντέλου, του αριθμού των χαρακτηριστικών εντός της περιοχής, και της ακρίβειας που του ταιριάζει.

2.2.3 Ο αλγόριθμος SIFT και τα 4 βήματα

Ο αλγόριθμος αποτελείται από τα ακόλουθα τέσσερα βήματα:

- i. Ανίχνευση ακρότατων στο χώρο της κλιμάκωσης (Scale-space extrema detection)
- ii. Εντοπισμός σημείων -κλειδιών (Patch Localization)
- iii. Ανάθεση προσανατολισμού (Orientation Assignment)
- iv. Περιγραφή των σημείων-κλειδιών (Patch Description)

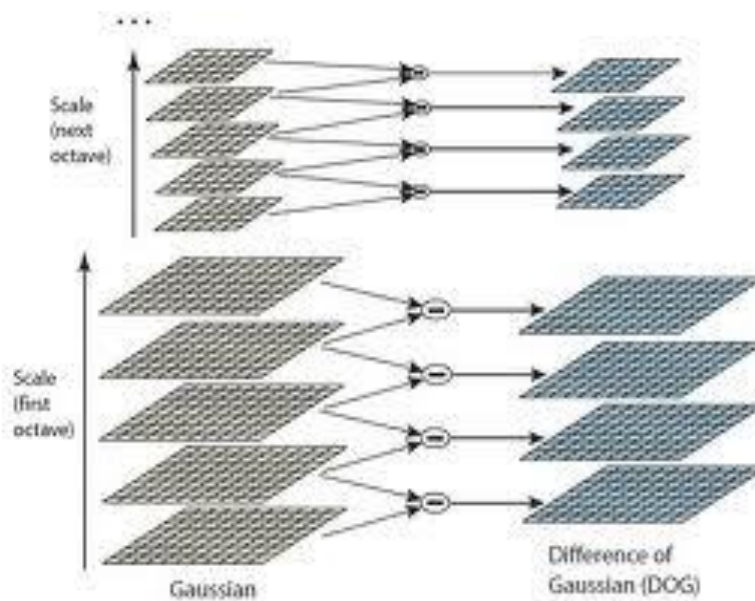
Ανίχνευση ακρότατων στο χώρο της κλιμάκωσης (Scale-space extrema detection)

Σε αυτό το βήμα εντοπίζονται τα σημεία ενδιαφέροντος, τα οποία ονομάζονται σημεία-κλειδιά όπως έχει προαναφερθεί. Για το σκοπό αυτό, η εικόνα υπόκειται σε συνέλιξη με Γκαουσιανά φίλτρα (Gaussian filters) σε διαφορετικές κλίμακες, και στη συνέχεια λαμβάνεται η διαφορά των διαδοχικών Gaussian-blurred (εικόνων με θόρυβο) εικόνων. Τα σημεία-κλειδιά λαμβάνονται στη συνέχεια ως μέγιστα / ελάχιστα της Διαφοράς του Gauss (Difference of Gaussians-DoG), που εμφανίζονται σε πολλαπλές κλίμακες. Συγκεκριμένα, μια εικόνα DoG $D(x,y,\sigma)$ δίνεται από:

$$D(x,y,\sigma) = L(x,y,k_i\sigma) - L(x,y, k_j\sigma) \quad (2.5)$$

Για να παραχθεί μια σειρά φωτογραφιών με διαφορετική κλίμακα από την αρχική εικόνα, εφαρμόζουμε το φίλτρο Gauss με ένα πολλαπλασιαστικό παράγοντα $k \geq 0$, $G(x,y,k\sigma)$ σε κλίμακα $k\sigma$ και παίρνουμε τη συνέλιξη του με την αρχική εικόνα $I(x,y)$, όπου $L(x,y,k\sigma)$ είναι η συνέλιξη της αρχικής εικόνας $I(x,y)$ οπότε προκύπτει:

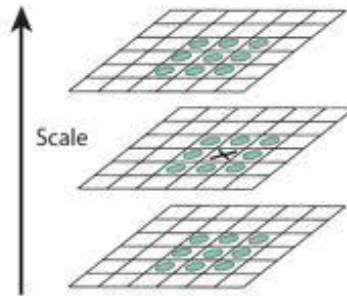
$$L(x,y,k\sigma) = G(x,y,k\sigma) * I(x,y) \quad (2.6)$$



Εικόνα 2.3 Αριστερά: Η αρχική εικόνα αλλάζει κλίμακα και γίνεται συνέλιξη των εικόνων που παράγονται με το φίλτρο Gauss. Δεξιά: Οι γειτονικές εικόνες αφαιρούνται και παράγεται η συνάρτηση διαφοράς του Gauss

Το επόμενο βήμα αφορά στον εντοπισμό του τοπικού ελαχίστου και μεγίστου της συνάρτησης διαφοράς του Gauss. Για να γίνει αυτό, συγκρίνουμε κάθε pixel της

εικόνας με τα 8 γειτονικά της, καθώς και με τα 26 κοντινότερα pixel των γειτονικών της εικόνων όπως φαίνεται στο παρακάτω σχήμα. Επιλέγεται το pixel του οποίου η συνάρτηση διαφορών του Gauss έχει τιμή μεγαλύτερη ή μικρότερη από τη συνάρτηση διαφορών του Gauss των υπολοίπων pixel με τα οποία συγκρίνεται.



Εικόνα 2.4 Το pixel που σημειώνεται με X συγκρίνεται με τα 8 γειτονικά του στην ίδια κλίμακα και με τα 18 pixels στις γειτονικές του εικόνες.

2.2.2 Εντοπισμός σημείων-κλειδιών

Σε αυτό το σημείο γίνεται ο εντοπισμός των σημείων-κλειδιών, όπου αρχικά χρησιμοποιείται για κάθε υποψήφιο σημείο-κλειδί η παρεμβολή των κοντινών σημείων. Η παρεμβολή υπολογίζεται με τετραγωνικό ανάπτυγμα Taylor της συνάρτησης Διαφοράς του Gauss (DoG) $D(x,y,\sigma)$ ως εξής:

$$D(x) = D + \frac{\partial D}{\partial x}x + \frac{1}{2}x^T \frac{\partial^2 D}{\partial x^2}x \quad (2.7)$$

Όπου D και οι παράγωγοί της υπολογίζονται για το υπόψηφιο σημείο-κλειδί και $x=(x,y,\sigma)$ είναι η απόσταση από το αυτό το σημείο. Η τοποθεσία του ακρότατου \hat{x} καθορίζεται υπολογίζοντας την παράγωγο ως προς x και θέτοντας την ίση μηδέν (0). Εάν η απόσταση του ακρότατου \hat{x} είναι μεγαλύτερη από 0.5 σε οποιαδήποτε διάσταση, αυτό είναι ένδειξη ότι το ακρότατο βρίσκεται πιο κοντά σε άλλο σημείο-κλειδί. Σε αυτήν την περίπτωση το υποψήφιο σημείο-κλειδί αλλάζει και η παρεμβολή υπολογίζεται στο σημείο αυτό. Ειδικά, η απόσταση

προστίθεται στο υποψήφιο σημείο-κλειδί για να πάρουμε την τοποθεσία του ακρότατου.

Για να απορρίψουμε τα σημεία-κλειδιά με μικρή αντίθεση υπολογίζεται το ανάπτυγμα Taylor δεύτερης τάξης στο ακρότατο \hat{x} . Αν η τιμή είναι μικρότερη από 0.03 τότε το σημείο-κλειδί απορρίπτεται. Αλλιώς διατηρείται με τελική θέση που δίνεται από τον τύπο:

$$y + \hat{x} \quad (2.8)$$

και κλίμακα σ , όπου y είναι η αρχική τοποθεσία του σημείου-κλειδιού σε κλίμακα σ .

Η συνάρτηση Διαφοράς του Gauss(DoG) έχει ισχυρές αντιδράσεις κατά μήκος των ακμών, ακόμη και αν το υποψήφιο σημείο-κλειδί δεν είναι ανθεκτικό σε μικρά επίπεδα θορύβου. Επομένως για να αυξηθεί η σταθερότητα πρέπει να περιοριστούν τα σημεία-κλειδιά τα οποία έχουν μεγάλη απόκριση, αλλά δεν έχει προσδιοριστεί επαρκώς η θέση τους(Wikipedia).

Για κακώς οριζόμενες κορυφές στη συνάρτηση Διαφοράς του Gauss(DoG), η κύρια καμπυλότητα σε ολόκληρη την ακμή θα είναι πολύ μεγαλύτερη από την κύρια καμπυλότητα κατά μήκος της. Η εύρεση αυτών των αρχικών κεφαλαίων καμπυλότητας για την επίλυση για τις ιδιοτιμές της δεύτερης τάξης της μήτρας Hessian(Hessian matrix) H:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.9)$$

Οι ιδιοτιμές του H είναι ανάλογες προς τις κύριες καμπυλότητες της D. Αποδεικνύεται ότι η αναλογία των δύο ιδιοτιμών είναι α η μεγαλύτερη και β η μικρότερη με λόγο $r = \alpha/\beta$. Άρα τα ίχνη του H, $D_{xx} + D_{yy}$ μας δίνουν το άθροισμα των δύο ιδιοτιμών.

2.2.3 Ανάθεση προσανατολισμού

Σε αυτό το βήμα, σε κάθε σημείο-κλειδί που έχει εντοπιστεί απαιτείται να ανατεθεί προσανατολισμός. Αυτό είναι το βασικό βήμα για την επίτευξη αναλλοίωτης περιστροφής, καθώς ο περιγραφέας του σημείου-κλειδιού μπορεί να αναπαρασταθεί σε σχέση με με αυτόν τον προσανατολισμό και ως εκ τούτου να επιτευχθεί αναλλοίωτη περιστροφή της εικόνας.

Αρχικά έχει ληφθεί μια Gaussian-smoothed εικόνα $L(x,y,\sigma)$ σε κλίμακα σ του σημείου-κλειδιού η οποία έχει ληφθεί έτσι ώστε όλοι οι υπολογισμοί να πραγματοποιούνται με έναν τρόπο αμετάβλητης-κλίμακας(scale-invariant). Δεδομένης μιας εικόνας $L(x,y)$ σε κλίμακα σ , το μέγεθος κλίσης $m(x,y)$ και ο προσανατολισμός $\theta(x,y)$ υπολογίζονται χρησιμοποιώντας τη διαφορά των pixel:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (2.10)$$

$$\theta(x,y) = \text{atan2}(L(x,y+1) - L(x,y-1), L(x+1,y) - L(x-1,y)) \quad (2.11)$$

Οι υπολογισμοί μεγέθους και κατεύθυνσης της κλίσης έχουν γίνει για κάθε pixel σε μια γειτονική περιοχή γύρω από το σημείο-κλειδί μιας Gaussian-blurred εικόνας L . Ένα ιστόγραμμα προσανατολισμού με 36 κάδους διαμορφώνεται, με κάθε κάδο να καλύπτει 10 μοίρες. Κάθε δείγμα στο γειτονικό παράθυρο προστίθεται σε ένα ιστόγραμμα κάδου σταθμίζεται από το μέγεθος κλίσης και με ένα Gaussian-weighted κυκλικό παράθυρο με κλίμακα σ να είναι 1.5 φορές μεγαλύτερη από το σημείο-κλειδί.

Τέλος, από το ιστόγραμμα επιλέγεται ο προσανατολισμός με την υψηλότερη συχνότητα, καθώς και οι προσανατολισμοί που αντιστοιχούν στο 80% της υψηλότερης συχνότητας και ανατίθενται στο σημείο κλειδί.

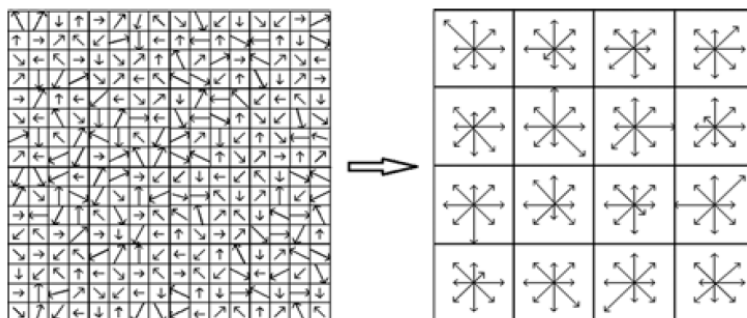
2.2.4 Περιγραφή των σημείων-κλειδιών

Στα προηγούμενα βήματα βρέθηκαν τοποθεσίες σημείων-κλειδιών σε συγκεκριμένες κλίμακες και τους ανατέθηκε ο προσανατολισμούς τους. Αυτό εξασφάλισε αναλλοίωτη την θέση της εικόνας, την κλίμακα και την περιστροφή της. Τώρα θέλουμε να υπολογίσουμε ένα διάνυσμα για κάθε σημείο-κλειδί το

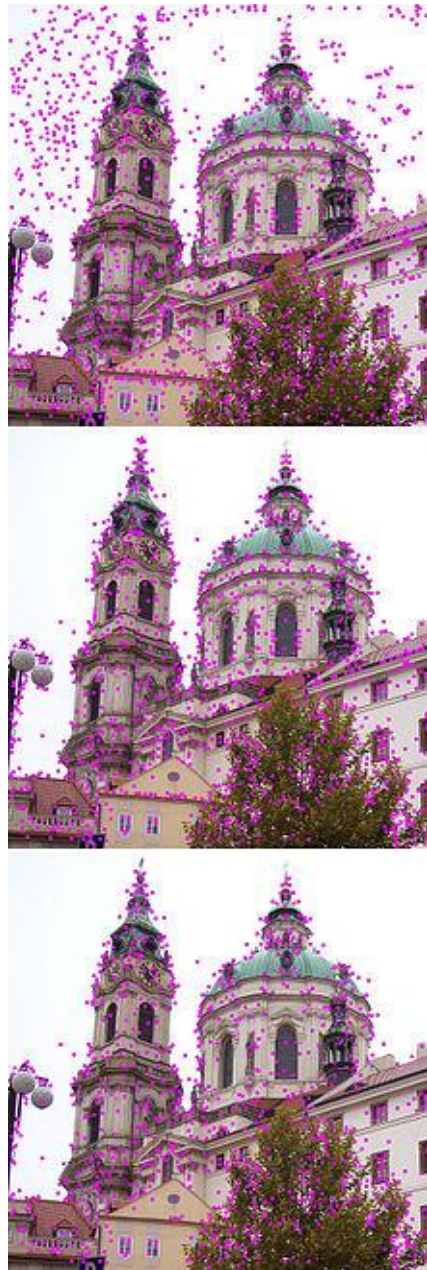
οποίο να περιλαμβάνει πληροφορία για την θέση, τον προσανατολισμό και την κλίμακά του.

Αρχικά δημιουργείται ένα σύνολο ιστογραμμάτων προσανατολισμού με γειτονίες 4x4 pixels έχοντας 8 κάδους η κάθεμία. Αυτά τα ιστογράμματα υπολογίζονται από το μέγεθος και τον προσανατολισμό των δειγμάτων σε μία περιοχή 16x16 γύρω από το σημείο-κλειδί έτσι ώστε κάθε ιστόγραμμα να περιέχει δείγματα από μία 4x4 υποπεριοχή της αρχικής περιοχής της γειτονιάς. Τα μεγέθη στη συνέχεια σταθμίζονται με μία Γκαουσιανή συνάρτηση (Gauss function) με κλίμακα σ ίση με το μισό του μεγέθους του παραθύρου. Ο περιγραφέας στη συνέχεια γίνεται διάνυσμα με όλες τα στοιχεία των ιστογραμμάτων. Δεδομένου ότι υπάρχουν $4 \times 4 = 16$ ιστογράμματα έχοντας το κάθε ένα 8 κάδους τότε το διάνυσμα μας αποτελείται από 128 στοιχεία. Αυτό το διάνυσμα κανονικοποιείται στη μονάδα μήκους έτσι ώστε να ενισχυθεί και να παραμείνει αναλλοίωτο σε affine αλλαγές του φωτισμού. Τέλος Για να μειωθούν οι επιπτώσεις του μη γραμμικού φωτισμού εφαρμόζεται ένα μέγιστο κατώφλι ίσο με 0.2 και έπειτα κανονικοποιείται ξανά.

Το τελικό στάδιο για τη συνταύτιση είναι η αντιστοίχιση των διανυσμάτων των περιγραφικών χαρακτηριστικών. Η διαδικασία αυτή πραγματοποιείται με την εύρεση του πλησιέστερου γείτονα, ο οποίος προκύπτει με υπολογισμό της ευκλείδειας απόστασης μεταξύ των διανυσμάτων και επιλογή εκείνου που δίνει τη μικρότερη τιμή. Η διαδικασία αυτή είναι απαιτητική υπολογιστικά και χρονοβόρα, για τον λόγο αυτό χρησιμοποιείται ο αλγόριθμος BBF, ο οποίος επιστρέφει τον κοντινότερο γείτονα με μεγάλη πιθανότητα και σταματά την αναζήτηση για αυτόν, ύστερα από έρευνα στους πρώτους 200 υποψηφίους κάθε σημείου



Εικόνα 2.5 Περιγραφή των σημείων-κλειδιών



Εικόνα 2.6 Πάνω εικόνα: Υστερα από την αρχική ανίχνευση,Μεσαία: Ο αλγόριθμος απορρίπτει τα σημεία-κλειδιά χαμηλής αντίθεσης, Κάτω: Τέλος φιλτράρονται οι άκρες

2.3 Αλγόριθμος SURF (Speeded-Up Robust Features)

Στην προηγούμενη ενότητα αναλύσαμε τον αλγόριθμο SIFT και τις λειτουργίες του. Ο SIFT όπως προαναφέραμε υλοποιεί έναν ανιχνευτή για τον προσδιορισμό σημείων ενδιαφέροντος σε μια εικόνα, πάνω στην ίδια λειτουργία βασίζεται και ο αλγόριθμος SURF (Speeded-Up Robust Features).

Η εύρεση αντιστοιχίσεων ανάμεσα σε δύο εικόνες της ίδιας σκηνής ή του ίδιου αντικειμένου είναι μέρος πολλών εφαρμογών υπολογιστικής όρασης. Η βαθμονόμηση κάμερας, 3D ανασυγκρότηση, καταγραφή εικόνας και αναγνώριση αντικειμένων είναι μερικά μόνο. Η αναζήτηση για τις διακριτές αντιστοιχίες μίας εικόνας - είναι ο στόχος αυτής της εργασίας - μπορούν να χωριστούν σε τρία βασικά στάδια. Πρώτον, τα «σημεία ενδιαφέροντος» που επιλέγονται σε διακριτικό περιοχές στην εικόνα, όπως γωνίες, blobs, και T-κόμβους(T-junctions)(16). Η πιο σημαντική ιδιότητα ενός ανιχνευτή σημείου ενδιαφέροντος είναι η επαναληψιμότητά του, δηλαδή αν βρίσκει αξιόπιστα τα ίδια σημεία ενδιαφέροντος κάτω από διαφορετικές συνθήκες προβολής. Στη συνέχεια, η γειτονιά κάθε σημείου ενδιαφέροντος αντιπροσωπεύεται από ένα διάνυσμα χαρακτηριστικών γνωρισμάτων. Αυτός ο περιγραφέας πρέπει να είναι διακριτικός και, ταυτόχρονα, ισχυρός στο θόρυβο, την ανίχνευση σφαλμάτων, γεωμετρικών και φωτομετρικών παραμορφώσεων. Τέλος, ο διανυσματικός περιγραφέας αντιστοιχίζεται μεταξύ διαφορετικών εικόνων. Η αντιστοίχιση αυτή βασίζεται σε μια απόσταση μεταξύ των διανυσμάτων, π.χ. Ευκλείδεια απόσταση. Η διάσταση του περιγραφέα έχει άμεσο αντίκτυπο στο χρόνο που αυτό απαιτεί και έχει μείωση του αριθμού των διαστάσεων.

Ο βασικός στόχος ο οποίος οδήγησε στη δημιουργία του SURF ήταν να αναπτυχθεί ένας ανιχνευτής και περιγραφέας ταυτόχρονα, ο οποίος συγκριτικά με το state-of-the-art είναι ταχύτερος στον υπολογισμό και δεν θα υστερεί σε απόδοση.

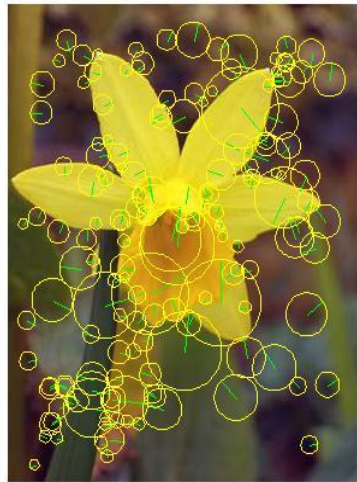
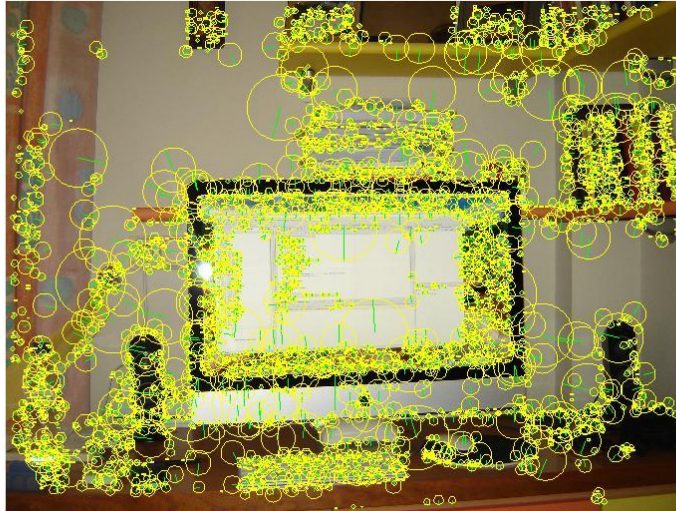
2.3.1 Τα σημεία ενδιαφέροντος

Ο πιο ευρέως διαδεδομένος ανιχνευτής είναι ο ανιχνευτής γωνιών Harris (Harris corner detector), ο οποίος προτάθηκε το 1988 και είναι βασισμένος στις ιδιοτιμές του πίνακα δεύτερης τάξης. Ωστόσο, οι γωνίες Harris δεν είναι σε σταθερή

κλίμακα. Ο Lindeberg αργότερα εισήγαγε την έννοια της αυτόματης επιλογής κλίμακας. Αυτό επιτρέπει τον εντοπισμό σημείων ενδιαφέροντος σε μια εικόνα, το καθένα με τη δική του χαρακτηριστική κλίμακα. Αυτός πειραματίστηκε με την ορίζουσα της μήτρας Hessian (Hessian matrix), καθώς και με τη λαπλασιανή μήτρα (η οποία αντιστοιχεί στα ίχνη της μήτρας Hessian) για την ανίχνευση δομών bloblike. Οι Mikolajczyk και Schmid εξευγένισαν αυτή τη μέθοδο, δημιουργώντας μία ισχυρή και σταθερή κλίμακα ανιχνευτών με υψηλή επαναληψιμότητα[29]. Χρησιμοποίησαν μια προσαρμοσμένη κλίμακα του μέτρου Harris και τον SURF για να επιλεγεί η κλίμακα. Εστιάζοντας στην ταχύτητα ο Lowe προσέγγισε τη LoG(Laplacian of Gaussian) με φίλτρο της DoG.

Αρκετοί άλλοι (σταθερής κλίμακας) ανιχνευτές σημείων ενδιαφέροντος είχαν προταθεί, όπως ο ανιχνευτής περιοχής ο οποίος προτάθηκε από τους Kadir και Brady και μεγιστοποιεί την εντροπία εντός της περιοχής. Μελετώντας τους ήδη υπάρχοντες ανιχνευτές και τις δημοσιευμένες συγκρίσεις που υπάρχουν συμπεραίνουμε ότι οι Hessian-based ανιχνευτές είναι πιο σταθεροί και επαναλαμβανόμενοι από ότι οι αντίστοιχοι Harris-based ανιχνευτές.

Στην εικόνα 2.7 φαίνονται δύο εικόνες με τα σημεία ενδιαφέροντος αποτυπωμένα με πράσινες τελείες. Η ακτίνα του κίτρινου κύκλου γύρω από κάθε σημείο είναι ανάλογη της κλίμακας στην οποία βρέθηκε.



Εικόνα 2.7 Δύο εικόνες και τα σημεία ενδιαφέροντός τους.

2.3.2 Προσδιορισμός των σημείων ενδιαφέροντος

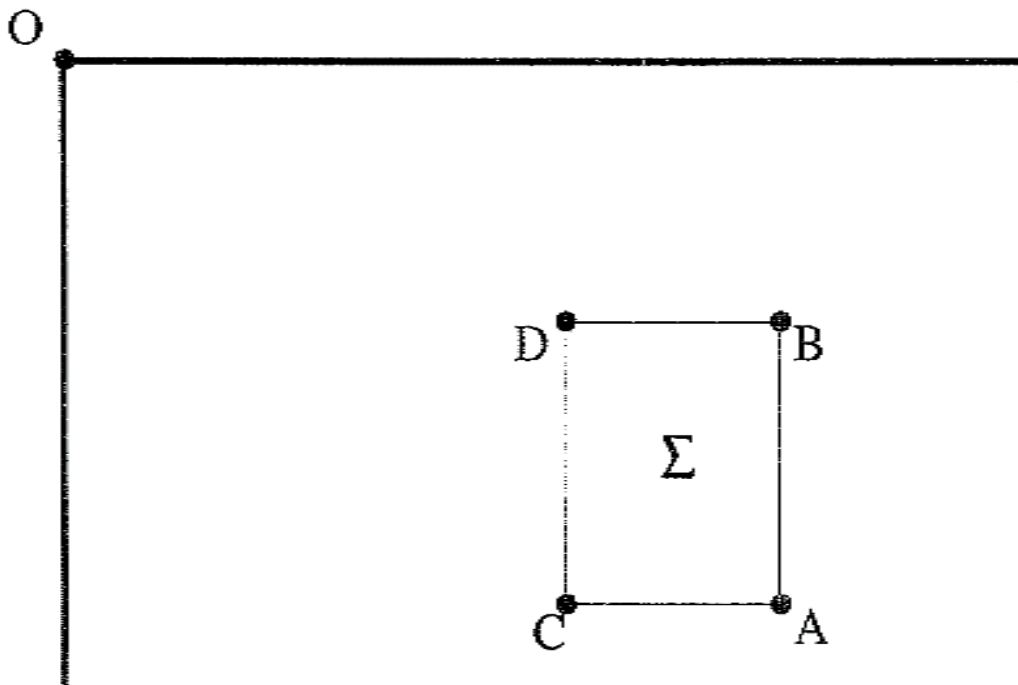
Για τον προσδιορισμό των σημείων ενδιαφέροντος η παραπάνω μέθοδος προτείνει μια ιδιαίτερα γρήγορη εκτίμηση της μήτρας Hessian. Η μήτρα Hessian είναι ο τετραγωνικός πίνακας των μερικών παραγώγων δεύτερης τάξης μιας συνάρτησης. Για να εκτιμηθεί η μήτρα, γίνεται χρήση των *integral images* με την χρήση των οποίων αυξάνεται μεγάλο μέρος της απόδοσης του SURF. Η *integral image* υπολογίζεται γρήγορα από μια εισαχθήσα εικόνα και χρησιμοποιείται έτσι ώστε να επιταχύνει τον υπολογισμό του κάθε ακραίου σημείου μιας ορθογώνιας

περιοχής. Δοσμένης μίας εικόνας I στη θέση $x=(x, y)$ η integral image I_Σ υπολογίζεται από το άθροισμα όλων των τιμών μεταξύ του σημείου και της προέλευσης δηλαδή:

$$I_\Sigma(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (2.12)$$

Χρησιμοποιώντας την integral image ο υπολογισμός της περιοχής ενός ακραίου σημείου της ορθογώνιας περιοχής απαιτεί μόνο τρεις αθροίσεις και τέσσερις προσβάσεις στη μνήμη για να υπολογιστεί το άθροισμα των φωτεινοτήτων σε οποιαδήποτε ορθογώνια περιοχή της εικόνας (2.8). Αν θεωρήσουμε ένα παραλληλόγραμμο που περικλείεται από τις κορυφές A, B, C και D τότε το άθροισμα των εντάσεων των pixels υπολογίζεται :

$$\Sigma = A - B - C + D \quad (2.13)$$



Εικόνα 2.8 Με χρήση των integral images το άθροισμα των φωτεινοτήτων στην περιοχή Σ υπολογίζεται με τρεις αθροίσεις και τέσσερις προσβάσεις στην μνήμη.

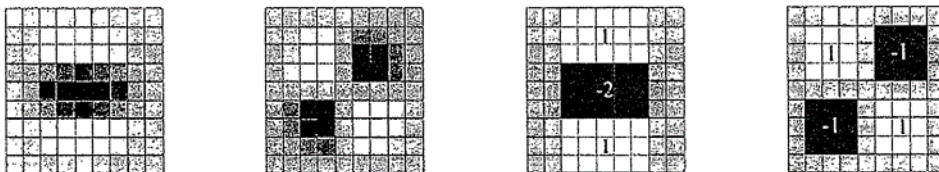
2.3.3 Η μήτρα Hessian

Ο SURF ανιχνευτής σημείων ενδιαφέροντος βασίζεται στην ορίζουσα της μήτρας Hessian. Προκειμένου να παρακινήσουμε τη χρήση της μήτρας, θεωρούμε μία συνεχή συνάρτηση δύο μεταβλητών. Δεδομένου σημείου $x = (x, y)$ σε μία εικόνα I , η μήτρα Hessian $H(x, \sigma)$ στο x σε κλίμακα σ ορίζεται ως εξής:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (2.11)$$

Όπου $L_{xx}(x, \sigma)$ είναι η συνέλιξη της γκαουσιανής μερικής παραγώγου δευτέρας τάξης (Gaussian second order derivative) $\frac{\partial^2}{\partial x^2} g(\sigma)$ με την εικόνα I στο σημείο x , ομοίως οι $L_{xy}(x, \sigma)$ και $L_{yy}(x, \sigma)$ είναι βέλτιστες για την ανάλυση κλίμακας-χώρου. Οι γκαουσιανές θεωρούνται βέλτιστες για την ανάλυση σε κλίμακες, αλλά μιας και πρέπει να διακριτοποιηθούν και να κατωφλιωθούν, αποκτούν αδυναμία στην επαναληψιμότητα τους σε περιστροφές γύρω από περιττά πολλαπλάσια του $\pi/4$. Αυτό συμβαίνει λόγω της τετραγωνικής φύσης του φίλτρου και ισχύει για όλους τους ανιχνευτές που χρησιμοποιούν τη μήτρα Hessian. Για τον ίδιο λόγο, η μέγιστη επαναληψιμότητα εμφανίζεται στα πολλαπλάσια των $\pi/2$ μοιρών. Για περαιτέρω απλούστευση των υπολογισμών οι μερικές παράγωγοι των γκαουσιανών προσεγγίζονται από φίλτρα-κουτιού (box filters) όπως φαίνεται στην εικόνα (2.9) χωρίς μείωση στην απόδοση.

Τα 9×9 φίλτρα της εικόνας (2.9) είναι οι προσεγγίσεις της γκαουσιανής μερικής παραγώγου δευτέρας τάξης με $\sigma=1.2$ και αντιπροσωπεύουν την χαμηλότερη κλίμακα υπολογισμού. Ας συμβολίσουμε τις προσεγγίσεις με D_{xx} , D_{yy} και D_{xy} .



Εικόνα 2.9 Από τα αριστερά: Οι διακριτοποιημένες γκαουσιανές μερικές παράγωγοι δευτέρας τάξης στην κατεύθυνση $-y$ και στην κατεύθυνση $-xy$ και οι δύο απλοποιήσεις τους.

Τα βάρη τα οποία εφαρμόζονται στις ορθογώνιες περιοχές διατηρούνται απλά για υπολογιστική αποδοτικότητα, αλλά πρέπει να ισορροπήσουν περαιτέρω στην έκφραση του παράγοντα Hessian με $\frac{|L_{xy}(1.2)|F|D_{xx}(1.2)|F}{|L_{xx}(1.2)|F|D_{xy}(1.2)|F} = 0.912 \dots \approx 0.9$. Αυτό μας δίνει :

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2.12)$$

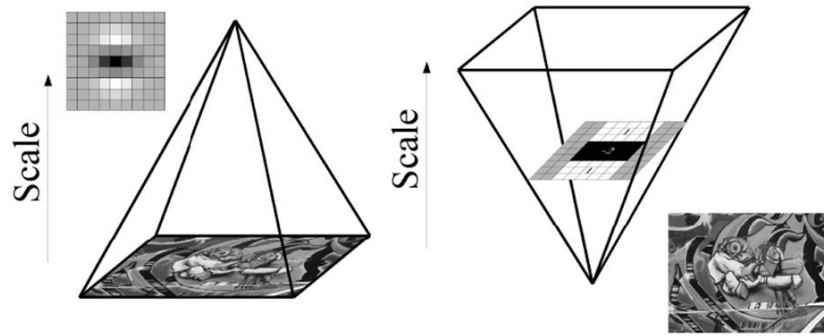
Επιπλέον, οι αποκρίσεις φίλτρου κανονικοποιούνται σε σχέση με το μέγεθος της μάσκας. Οι κλίμακες χώρου εφαρμόζονται συνήθως σαν πυραμίδες. Οι εικόνες είναι επανειλημμένα εξομαλυμένες προκειμένου να φθάσουν το υψηλότερο επίπεδο ης πυραμίδας. Λόγω της χρήσης φίλτρων-κουτιά και των integral images δεν είναι ανάγκη να εφαρμόζουμε κατ'επανάληψη το ίδιο φίλτρο στην έξοδο ενός προηγούμενως φιλτραρισμένου στρώματος, αλλά μπορούμε να εφαρμόσουμε φίλτρα φίλτρα οποιουδήποτε μεγέθους της ίδιας ταχύτητας στην αρχική εικόνα.

Το αποτέλεσμα των παραπάνω είναι ένα 9x9 φίλτρο το οποίο θεωρείται το αρχικό στρώμα της κλίμακας σε κλίμακα $\sigma = 1.2$. Τα ακόλουθα στρώματα λαμβάνονται με φιλτράρισμα της εικόνας εφαρμόζοντας σταδιακά μεγαλύτερες μάσκες. Συγκεκριμένα αυτό έχει ως αποτέλεσμα τα φίλτρα μεγέθους 9x9, 15x15, 21x21, 27x27, κλπ.

2.3.4 Κατασκευή της κλίμακας χώρου

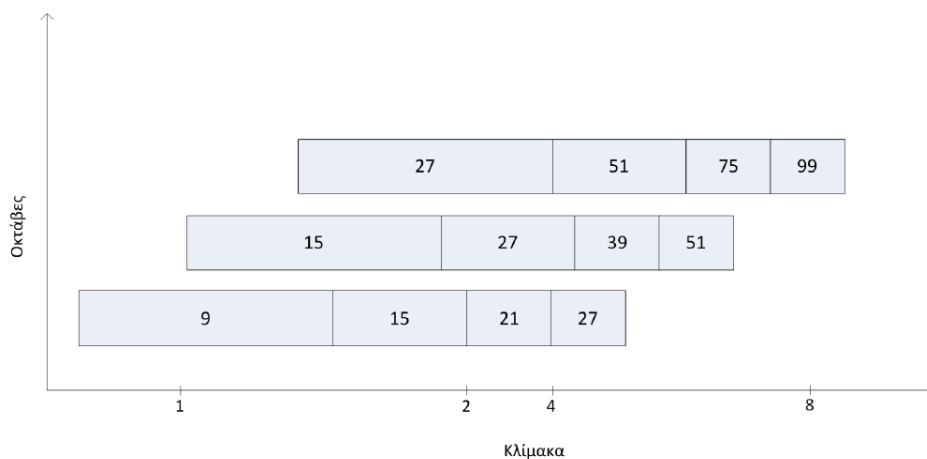
Για την ανίχνευση των σημείων ενδιαφέροντος χρησιμοποιώντας την ορίζουσα Hessian είναι απαραίτητο να εισαγάγουμε την έννοια της κλίμακας χώρου(scale space)(έχει προαναφερθεί). Μια κλίμακα χώρου είναι μια συνεχής συνάρτηση η οποία μπορεί να χρησιμοποιείται για να βρεθούν τα ακρότατα σε όλες τις πιθανές κλίμακες. Στην τεχνητή όραση η κλίμακα χώρου συνήθως υλοποιείται ως μια πυραμίδα, όπου η εικόνα εισόδου συνελίσεται κατ'επανάληψη με Gaussian kernel και επανειλημμένη υπο-δειγματοληψία (μειωμένου μεγέθους). Η μέθοδος αυτή χρησιμοποιείται σε μεγάλο βαθμό στον SIFT, αλλά δεδομένου ότι κάθε στρώμα εξαρτάται από το προηγούμενο, και οι εικόνες πρέπει να αλλάζουν το μέγεθός τους δεν είναι υπολογιστικά εφικτό(15).

Αυτό επιτρέπει την ταυτόχρονη επεξεργασία των πολλαπλών στρώσεων της κλίμακας χώρου πυραμίδας. Στην εικόνα (2.10) απεικονίζεται η διαφορά μεταξύ της κλασσικής δομής κλίμακας χώρου και της SURF προσέγγισης.



Εικόνα 2.10 Filter Pyramid: Η παραδοσιακή προσέγγιση για την κατασκευή μιας κλίμακας χώρου (αριστερά). Η προσέγγιση SURF (δεξιά) αφήνει την αρχική εικόνα αμετάβλητη και ποικίλλει μόνο το μέγεθος του φίλτρου.

Για τα SURF αντί να γίνεται υποδειγματοληψία της εικόνας σε κάθε επίπεδο, μεγαλώνει το μέγεθος του τετραγωνικού φίλτρου με το οποίο αυτή απλοποιείται. Αυτό κάνει την εξαγωγή ταχύτερη καθώς με τις integral images ο χρόνος εφαρμογής ενός φίλτρου στην εικόνα είναι ανεξάρτητος από το μέγεθος του φίλτρου. Έτσι η εικόνα φιλτράρεται αρχικά με τα 9×9 φίλτρα ως ελάχιστη κλίμακα και έπειτα διαδοχικά με φίλτρα μεγαλύτερου μεγέθους, αυξάνοντας κάθε φορά το παράθυρο ανά τουλάχιστον ένα pixel σε κάθε πλευρά της μάσκας που οδηγεί συνολική αύξηση 6 pixel σε κάθε επίπεδο για ένα φίλτρο. Για την δεύτερη κλίμακα η εικόνα θα φιλτραριστεί με φίλτρο μεγέθους 15×15 pixel. Η διακριτή φύση των integral images επιβάλλει αυτόν τον κβαντισμό των κλιμάκων.



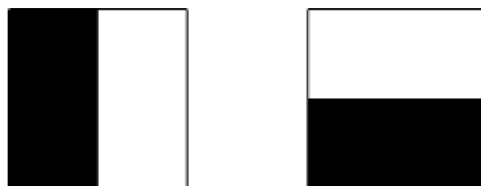
Εικόνα 2.11 Σχηματική παρουσίαση του μεγέθους των τετραγωνικών φίλτρων για τις τρεις πρώτες οκτάβες

Για να καθοριστεί η ακριβής θέση του σημείου ενδιαφέροντος πάνω στην εικόνα και ανεξάρτητα από την κλίμακα, εκτελείται ένας αλγόριθμος για τον εντοπισμό τοπικών μεγίστων ανάμεσα σε κλίμακες σε μία $3 \times 3 \times 3$ (κλίμακες) γειτονιά όπως προτείνουν οι Neubeck και Van Gool. Έπειτα, τα τοπικά μέγιστα από τις ορίζουσες της μήτρας υπόκεινται σε παρεμβολή με τη μέθοδο που είχε προτείνει ο Lowe (15), διαδικασία ιδιαίτερα σημαντική γιατί σε αρκετές περιπτώσεις οι μεταβολές της κλίμακας σε διαδοχικά επίπεδα είναι αρκετά μεγάλες.

Σύμφωνα με τα παραπάνω, η μικρότερη δυνατή κλίμακα, μετά και την παρεμβολή είναι $\sigma = 1.6 = 1.2 * 12/9$ που αντιστοιχεί σε φίλτρο 12×12 και η μεγαλύτερη $\sigma = 3.2 = 1.2 * 24/9$.

2.3.4 Περιγραφέας σημείων ενδιαφέροντος

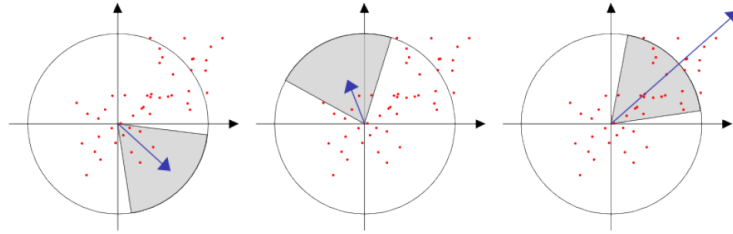
Ο περιγραφέας των SURF περιγράφει την κατανομή των pixels εντάσεων στην περιοχή γύρω από το σημείο ενδιαφέροντος, παρόμοια με την πληροφορία για τις κατευθύνσεις των παραγώγων που εξάγεται από τα SIFT. Αντί για την κατανομή των παραγώγων όμως, στα SURF χρησιμοποιούνται οι κατανομές των κυματιδίων Haar πρώτης τάξης (first order Haar wavelet responses) στις κατευθύνσεις x και y . Η ταχύτητα εξαγωγής των περιγραφέων αυξάνεται πολύ με την χρήση των integral images και για να κρατηθεί μικρό το μέγεθος του περιγραφέα εν τέλει χρησιμοποιούνται 64 διαστάσεις.



Εικόνα 2.12 Κυματίδια Haar (Haar wavelets) Το αριστερό φίλτρο υπολογίζει την απόκριση στην x κατεύθυνση και το δεξιό στην y κατεύθυνση

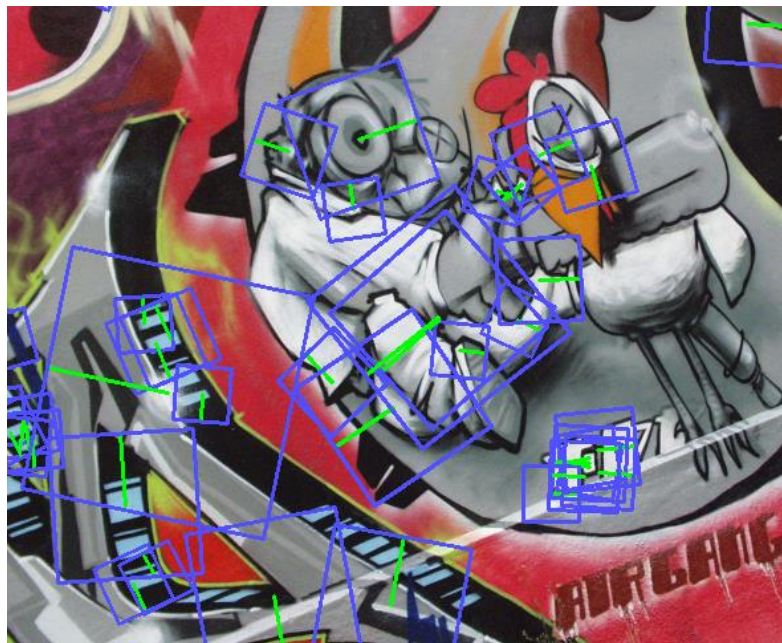
Για να υπάρχει ανεξαρτησία στην περιστροφή μιας εικόνας, καθορίζεται σε κάθε σημείο ενδιαφέροντος ένας κύριος προσανατολισμός. Αρχικά υπολογίζονται οι αποκρίσεις των κυματιδίων Haar στις κατευθύνσεις x και y σε μια ακτίνα 6σ γύρω από το σημείο με σ να είναι η κλίμακα στην οποία βρέθηκε το σημείο. Το μέγεθος των κυματιδίων είναι επίσης εξαρτώμενο από την κλίμακα, καθώς έχουν μέγεθος 4σ . Ως κύριο προσανατολισμό ορίζουμε το μακρύτερο διάνυσμα που προκύπτει από την άθροιση των αποκρίσεων σε ένα κυλιόμενο παράθυρο τόξου $\pi/3$. Σε κάθε

θήση τα x και y εντός του τμήματος αθροίζονται και χρησιμοποιούνται για να σχηματίσουν ένα νέο διάνυσμα. Το μακρύτερο διάνυσμα προσδίδει τον προσανατολισμό του, το σημείο ενδιαφέροντος. Η διαδικασία αυτή απεικονίζεται παρακάτω:



Εικόνα 2.13 Εντοπισμός του κύριου προσανατολισμού από το κυλιόμενο παράθυρο.

Το πρώτο βήμα για την εξαγωγή του SURF περιγραφέα είναι να κατασκευαστεί ένα τετραγωνικό παράθυρο γύρω από το σημείο ενδιαφέροντος. Αυτό το παράθυρο περιέχει τα pixels που θα αποτελέσουν εισαγωγές στον περιγραφέα και έχει μέγεθος 20σ. Επιπλέον, το παράθυρο είναι προσανατολισμένο προς την κατεύθυνση που βρέθηκαν στο προηγούμενο κεφάλαιο, έτσι ώστε όλοι οι επόμενοι υπολογισμοί να είναι σε σχέση προς αυτή την κατεύθυνση.



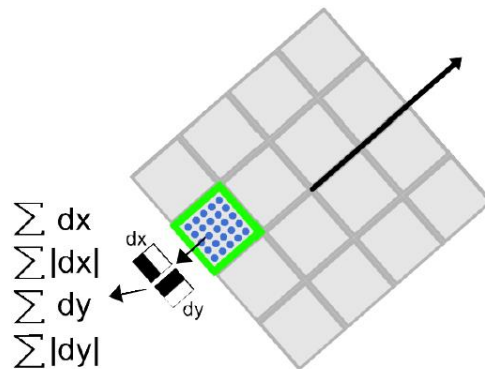
Εικόνα 2.14 Το μέγεθος του παραθύρου είναι 20 φορές το μέγεθος του ανιχνευόμενου σημείου και είναι προσανατολισμένο κατά μήκος της κυρίαρχης κατεύθυνσης που φαίνεται με πράσινο (opensurf).

Ο περιγραφέας παραθύρου χωρίζεται σε υποπεριοχές των 4x4. Μέσα σε κάθε μία από αυτές τις υποπεριοχές, τα Haar wavelets μεγέθους 2σ υπολογίζονται για 25 κατανεμημένα σημεία δείγματος. Αν αναφερόμαστε στις x και y αποκρίσεις κυματιδίων από dx και dy αντιστοίχως κατόπιν για αυτά τα 25 σημεία δειγματοληψίας (δηλαδή κάθε υποπεριοχής), συλλέγουμε,

$$u_{\text{υποπεριοχή}} = \left[\sum dx, \sum dx, \sum |dx|, \sum |dy| \right] \quad (2.13)$$

Ως εκ τούτου κάθε υποπεριοχή συνεισφέρει τέσσερις τιμές στο διάνυσμα του περιγραφέα το οποίο οδηγεί σε ένα συνολικό διάνυσμα μήκους $4 \times 4 \times 4 = 64$. Ο προκύπτων SURF περιγραφέας είναι αναλλοίωτος σε περιστροφή, κλίμακα, φωτεινότητα και μετά την αναγωγή προς μονάδα μήκους, αντίθεση.

Για να υπάρχει ανεξαρτησία από την αντίθεση (contrast) μετατρέπουμε το διάνυσμα του περιγραφέα σε μοναδιαίο.



Εικόνα 2.15 Το πράσινο τετράγωνο δεσμεύει μία από τις 16 υποπεριοχές, όπως απεικονίζεται οι x και y αποκρίσεις υπολογίζονται σε σχέση με το κυρίαρχο προσανατολισμό.

Κεφάλαιο 3

Αλγόριθμος RANSAC (Random Sample Consensus)

3.1 Εισαγωγή

Ο αλγόριθμος RANSAC είναι η συντομογραφία του «Random Sample Consensus». Αρχικά πρέπει να αναφερθεί πως αυτός ο αλγόριθμος είναι μια επαναληπτική μέθοδος για την εκτίμηση των παραμέτρων ενός μαθηματικού μοντέλου από ένα σύνολο παρατηρούμενων δεδομένων το οποίο περιέχει outliers. Ο RANSAC είναι ένας μη αιτιοκρατικός αλγόριθμος υπό την έννοια ότι παράγει ένα λογικό αποτέλεσμα μόνο με μια ορισμένη πιθανότητα, με αυτήν την πιθανότητα να αυξάνεται με τον αριθμό των επαναλήψεων. Ο RANSAC δημοσιεύτηκε για πρώτη φορά από τους κ.Fischler και κ.Bolles στο SRI International το 1981 [17]. Επίσης ο αλγόριθμος RANSAC έχει χρησιμοποιηθεί ευρέως σε διάφορα επιστημονικά και μη πεδία, όπως το πεδίο της μηχανικής αλλά και της τεχνητής όρασης.

Μία βασική παραδοχή είναι ότι τα δεδομένα αποτελούνται από 'inliers', δηλαδή τα δεδομένα των οποίων η κατανομή τους μπορεί να εξηγηθεί από ένα συγκεκριμένο σύνολο των παραμέτρων ενός μοντέλου, αν και μπορεί τα δεδομένα να έχουν πειραχθεί από θόρυβο και από 'outliers' τα οποία είναι δεδομένα τα οποία δεν ταιριάζουν με το μοντέλο. Τα 'outliers' υπάρχει πιθανότητα να έχουν προσέλθει, π.χ. από τα outlier του θορύβου ή από εσφαλμένες μετρήσεις ή λανθασμένες υποθέσεις σχετικά με την ερμηνεία των δεδομένων. Ο RANSAC τέλος υποθέτει ότι λαμβάνοντας υπόψη ένα (συνήθως μικρό) σύνολο 'inliers', υπάρχει μια διαδικασία η οποία μπορεί να εκτιμήσει τις παραμέτρους ενός μοντέλου που επεξηγεί άριστα ή ταιριάζει με αυτά τα δεδομένα.

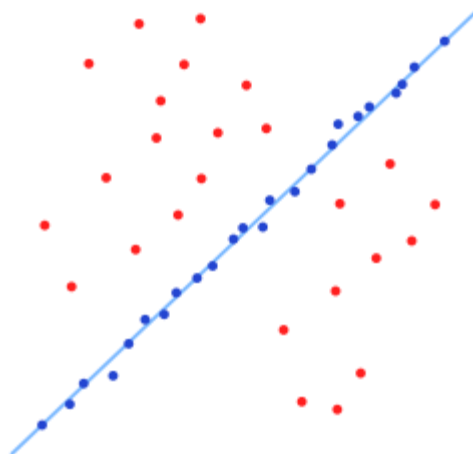
3.2 Γενική επισκόπηση αλγορίθμου

Για να γίνει πιο κατανοητός ο αλγόριθμος και στη συνέχεια να αναλύσουμε τις υπόλοιπες πτυχές του δίνεται παρακάτω ένα απλό παράδειγμα. Ένα τέτοιο παράδειγμα είναι η τοποθέτηση μιας γραμμής σε δύο διαστάσεις σε ένα σύνολο παρατηρήσεων. Υποθέτοντας ότι αυτό το σύνολο περιέχει δύο inliers, δηλαδή, τα σημεία που κατά προσέγγιση μπορεί να τοποθετηθούν σε μία γραμμή, και outliers, τα σημεία που δεν μπορούν να τοποθετηθούν σε αυτή τη γραμμή. Μια απλή μέθοδος των ελαχίστων τετραγώνων για την τοποθέτηση της γραμμής γενικά θα παράγει μια σειρά, με μια κακή εφαρμογή στα inliers. Ο λόγος είναι ότι έχουν τοποθετηθεί με βέλτιστο τρόπο σε όλα τα σημεία, συμπεριλαμβανομένων των outliers. Ο RANSAC, από την άλλη, μπορεί να παράγει ένα μοντέλο το οποίο υπολογίζεται μόνο από inliers, υπό την προϋπόθεση ότι η πιθανότητα της επιλογής μόνο inliers κατά την επιλογή των δεδομένων είναι επαρκώς υψηλή. Δεν υπάρχει "εγγύηση" για την κατάσταση αυτή, ωστόσο, υπάρχει ένας αριθμός

παραμέτρων αλγορίθμου που πρέπει να επιλέγονται προσεκτικά για να διατηρήσει το επίπεδο της πιθανότητας εύλογα υψηλό.



Εικόνα 3.1 Ένα σύνολο δεδομένων με πολλά outliers για τα οποία πρέπει να τοποθετηθεί μια γραμμή.



Εικόνα 3.2 Μετά την τοποθέτηση της γραμμής σύμφωνα με τον RANSAC. Τα outliers δεν έχουν καμία επίδραση στο αποτέλεσμα.

Σαν είσοδο ο αλγόριθμος RANSAC λαμβάνει ένα σύνολο παρατηρούμενων δεδομένων, ένα παραμετροποιημένο μοντέλο το οποίο μπορεί να εξηγήσει ή να τοποθετηθεί με τις παρατηρήσεις, και ορισμένες παράμετροι εμπιστοσύνης. Ο RANSAC επιτυγχάνει το στόχο του επιλέγοντας επαναληπτικά ένα τυχαίο υποσύνολο των αρχικών δεδομένων. Αυτά τα δεδομένα είναι υποθετικά inliers και αυτή η υπόθεση στη συνέχεια δοκιμάζεται με την ακόλουθη διαδικασία:

1. Ένα μοντέλο αντιζητιχίζεται με τα υποθετικά inliers, δηλαδή όλες οι ελεύθεροι παράμετροι του μοντέλου ανακατασκευάζονται από τα inliers.
2. Όλα τα άλλα δεδομένα, στη συνέχεια δοκιμάζονται έναντι του προσαρμοσμένου μοντέλου και αν ένα σημείο ταιριάζει καλά με το εκτιμώμενο μοντέλο, θεωρείται ένα υποθετικό inlier.
3. Το εκτιμώμενο μοντέλο είναι αρκετά καλό, αν έχουν ταξινομηθεί πολλά σημεία ως υποθετικά inliers.
4. Το μοντέλο επαναυπολογίζεται από όλα τα υποθετικά inliers, διότι είχε εκτιμηθεί μόνο από το αρχικό σύνολο υποθετικών inliers.
5. Τέλος, το μοντέλο αξιολογείται με εκτίμηση σφάλματος των inliers σε σχέση με το μοντέλο.

Αυτή η διαδικασία επαναλαμβάνεται με έναν σταθερό αριθμό, κάθε φορά που παράγεται, είτε ένα μοντέλο το οποίο απορρίπτεται επειδή είναι πολύ λίγα τα σημεία (ταξινομούνται ως inliers) είτε ένα εκλεπτυσμένο μοντέλο μαζί με ένα αντίστοιχο μέτρο σφάλματος. Στην τελευταία περίπτωση, κρατάμε το εκλεπτυσμένο μοντέλο, αν το σφάλμα είναι μικρότερο από το τελευταίο αποθηκευμένο μοντέλο.

Ο αλγόριθμος RANSAC σε ψευδοκώδικα λειτουργεί ως εξής (Wikipedia):

Input:

data - a set of observations

model - a model that can be fitted to data

n - the minimum number of data required to fit the model

k - the number of iterations performed by the algorithm

t - a threshold value for determining when a datum fits a model

d - the number of close data values required to assert that a model fits well to data

output:

best_model - model parameters which best fit the data (or nil if no good model is found)

best_consensus_set - data points from which this model has been estimated

best_error - the error of this model relative to the data

iterations := 0

best_model := nil

best_consensus_set := nil


```
best_error := infinity
while iterations < k
  maybe_inliers := n randomly selected values from data
  maybe_model := model parameters fitted to maybe_inliers
  consensus_set := maybe_inliers

  for every point in data not in maybe_inliers
    if point fits maybe_model with an error smaller than t
      add point to consensus_set

  if the number of elements in consensus_set is > d
    (this implies that we may have found a good model,
    now test how good it is)
    this_model := model parameters fitted to all points in consensus_set
    this_error := a measure of how well this_model fits these points
    if this_error < best_error
      (we have found a model which is better than any of the previous ones,
      keep it until a better one is found)
      best_model := this_model
      best_consensus_set := consensus_set
      best_error := this_error
    increment iterations
return best_model, best_consensus_set, best_error
```

3.3 Παράμετροι του αλγορίθμου

Οι τιμές των παραμέτρων t και d πρέπει να καθορίζονται από τις ειδικές απαιτήσεις που σχετίζονται με την εφαρμογή και το σύνολο δεδομένων, ενδεχομένως βασισμένες σε κάποια πειραματική αξιολόγηση. Η παράμετρος k (αριθμός επαναλήψεων), ωστόσο μπορεί να προσδιοριστεί από ένα θεωρητικό αποτέλεσμα. Έστω, p ορίζεται η πιθανότητα ο αλγόριθμος RANSAC να επιλέξει μόνο inliers σε κάποια επανάληψη, από τα δεδομένα εισόδου όταν ο αλγόριθμος επιλέγει n σημεία από τα οποία εκτιμώνται οι παράμετροι του μοντέλου. Όταν συμβαίνει αυτό, το προκύπτον μοντέλο είναι πιθανόν να είναι χρήσιμο, έτσι τελικά p δίνεται η πιθανότητα ο αλγόριθμος να παράγει ένα χρήσιμο αποτέλεσμα.

Έστω w η πιθανότητα να επιλέξουμε ένα inlier κάθε φορά που έχει επιλεγεί ένα ενιαίο σημείο, δηλαδή,

$w = \text{αριθμός inliers στα δεδομένα} / \text{αριθμός σημείων στα δεδομένα}.$

Μια κοινή περίπτωση είναι ότι το w δεν είναι γνωστό εκ των προτέρων, αλλά μπορεί να δοθεί κάποια αξία. Ας υποθέσουμε τώρα ότι τα n σημεία που απαιτούνται για την εκτίμηση ενός μοντέλου που επιλέγονται ανεξάρτητα, w^n είναι η πιθανότητα ότι όλα τα n σημεία να είναι inliers και $1-w^n$ η πιθανότητα ότι τουλάχιστον ένα από τα n σημεία είναι outlier, μια υπόθεση που υποδηλώνει ότι ένα κακό μοντέλο θα εκτιμηθεί από αυτό το σύνολο σημείων. Αυτή η πιθανότητα για τη δύναμη του k είναι η πιθανότητα ότι ποτέ ο αλγόριθμος δεν επιλέγει ένα σύνολο n σημείων τα οποία είναι όλα inliers και αυτό πρέπει να είναι $1-p$. Κατά συνέπεια [18][19];

$$1-p = (1 - w^n)^k \quad (3.1)$$

η οποία, μετά τη λογαρίθμηση και των δύο πλευρών, οδηγεί σε:

$$k = \frac{\log(1-p)}{\log(1-w^n)} \quad (3.2)$$

Αυτό το αποτέλεσμα, προϋποθέτει ότι τα n σημεία δεδομένων που επιλέγονται ανεξάρτητα, δηλαδή, ένα σημείο που έχει επιλεγεί, αντικαθίσταται μία φορά και μπορεί να επιλεγεί και πάλι στην ίδια επανάληψη. Αυτή δεν είναι μια λογική προσέγγιση και η προκύπτουσα τιμή για το k πρέπει να ληφθεί ως ένα ανώτερο

όριο στην περίπτωση που τα σημεία επιλέγονται χωρίς αντικατάσταση. Για παράδειγμα, στην περίπτωση της εύρεσης μιας γραμμής η οποία ταιριάζει με τα δεδομένα που απεικονίζονται στο παραπάνω σχήμα, ο αλγόριθμος RANSAC επιλέγει τυπικά 2 σημεία σε κάθε επανάληψη και υπολογίζει `maybe_model` ως η διαχωριστική γραμμή μεταξύ των σημείων και τότε είναι το κρίσιμο σημείο καθώς τα δύο σημεία είναι διακριτά.

Τέλος η τυπική απόκλιση (standard deviation) η πολλαπλάσια αυτής μπορούν να προστεθούν στο k . Η τυπική απόκλιση του k ορίζεται ως :

$$SD(k) = \frac{\sqrt{1-w^n}}{w^n} \quad (3.3)$$

3.4 Πλεονεκτήματα και μειονεκτήματα

Θα ήταν αρκετά χρήσιμο να παραθέσουμε τα πλεονεκτήματα και μειονεκτήματα του αλγόριθμου RANSAC καθώς μπορεί να γίνει ακόμη πιο κατανοητός.

Ένα πλεονέκτημα του RANSAC είναι η ικανότητά του να κάνει «σθεναρή» εκτίμηση των παραμέτρων ενός μοντέλου, δηλαδή, μπορεί να εκτιμήσει τις παραμέτρους με αρκετά υψηλό βαθμό ακρίβειας, ακόμα και όταν ένας σημαντικός αριθμός των outliers είναι παρών στο σύνολο δεδομένων.

Από την άλλη ένα μειονέκτημα του RANSAC είναι ότι δεν υπάρχει άνω όριο για το χρόνο που χρειάζεται για να υπολογίσει αυτές τις παραμέτρους. Όταν ο αριθμός των επαναλήψεων που έχουν υπολογιστεί είναι περιορισμένος, τότε η ληφθήσα λύση μπορεί να μην είναι η βέλτιστη, και ακόμη δεν μπορεί να είναι μία λύση που να ταιριάζει τα δεδομένα με ένα καλό τρόπο. Με τον τρόπο ο RANSAC προσφέρει ένα trade-off (αυξομείωση). Υπολογίζοντας ένα μεγαλύτερο αριθμό επαναλήψεων η πιθανότητα ενός παραγόμενου λογικού μοντέλου αυξάνεται. Ένα άλλο μειονέκτημα του RANSAC είναι ότι απαιτεί τη ρύθμιση της επίλυσης συγκεκριμένων ορίων.

Ο RANSAC μπορεί να εκτιμήσει μόνο ένα μοντέλο για ένα συγκεκριμένο σύνολο δεδομένων. Όπως για κάθε ένα μοντέλο προσέγγισης, όταν υπάρχουν δύο (ή περισσότερες) περιπτώσεις μοντέλων, επίσης ο RANSAC μπορεί να αποτύχει να μην βρει ούτε ένα. Ο μετασχηματισμός Hough είναι μια εναλλακτική τεχνική εκτίμησης που μπορεί να είναι χρήσιμη όταν ψρουζιάζονται περισσότερα από ένα παράδειγματα μοντέλων.

ΚΕΦΑΛΑΙΟ 4

Μεθοδολογία Εφαρμογής

4.1 Βασικές Λειτουργίες

Σε αυτό το κεφάλαιο θα γίνει μια περιγραφή των βασικών λειτουργιών του προγράμματος που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής. Η περιγραφή θα ξεκινήσει. Η ανάλυση θα ξεκινήσει από το πρώτο τμήμα του προβλήματος που είναι η αντιστοίχιση σημείων μεταξύ δύο εικόνων και θα περάσει στη συνέχεια στις υπόλοιπες ενέργειες όπως η συρραφή των εικόνων.

4.2 Αντιστοίχιση σημείων

Αφού έχουμε εξάγει τα χαρακτηριστικά με μεθόδους οι οποίες έχουν αναφερθεί παραπάνω κατόπιν αντιστοιχίζουμε τα κοινά σημεία των δύο εικόνων.

Σε αυτό το στάδιο ο στόχος μας είναι να βρεθούν όλες οι αντιστοιχίσεις μεταξύ των εικόνων. Τα ενωμένα σύνολα των αντιστοιχίσεων των εικόνων γίνονται αργότερα πανοράματα(panoramas). Δεδομένου ότι κάθε εικόνα θα μπορούσε ενδεχομένως να ταιριάζει με οποιαδήποτε άλλη, το πρόβλημα αυτό εμφανίζεται αρχικά τετραγωνικά ως προς τον αριθμό των εικόνων. Ωστόσο, είναι απαραίτητο να αντιστοιχεί κάθε εικόνα σε έναν μικρό αριθμό επικαλυπτόμενων εικόνων έτσι ώστε να λάβει μία καλή λύση για την γεωμετρία της εικόνας.

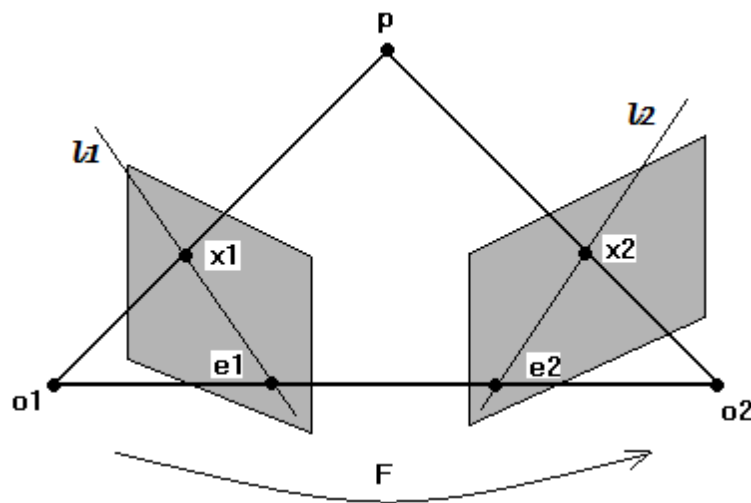
Από το βήμα της αντιστοίχισης των χαρακτηριστικών(SURF), έχουμε εντοπίσει τις εικόνες οι οποίες έχουν έναν μεγάλο αριθμό αντιστοιχίσεων μεταξύ τους. Θεωρούμε έναν σταθερό αριθμό εικόνων m , που έχουν τον μεγαλύτερο αριθμό αντιστοιχίσεων των χαρακτηριστικών για την τρέχουσα εικόνα, ως πιθανό αριθμό των αντιστοιχίσεων των εικόνων (χρησιμοποιούμε $m = 6$). Πρώτον, χρησιμοποιούμε τον RANSAC για να επιλέξει μια σειρά από inliers που είναι

συμβατά με την ομογραφία(homography) μεταξύ των εικόνων. Στη συνέχεια εφαρμόζουμε ένα μοντέλο πιθανοτήτων για να ελέγξει την αντιστοίχιση.

4.2.1 Υπολογισμός της ομογραφίας(homography)

Όπως αναφέρθηκε και παραπάνω ο αλγόριθμος RANSAC έχει σαν στόχο την εύρεση της ομογραφίας ανάμεσα σε δύο εικόνες, δηλαδή τον προοπτικό μετασχηματισμό ο οποίος μετασχηματίζει κάθε σημείο της μίας εικόνας με κάθε σημείο x_1 της μίας εικόνας σε κάθε σημείο x_2 της άλλης στην περίπτωση που υπάρχουν πολλές λάθος αντιστοιχίσεις.

$$x_2 = x_1 H \quad (4.1)$$



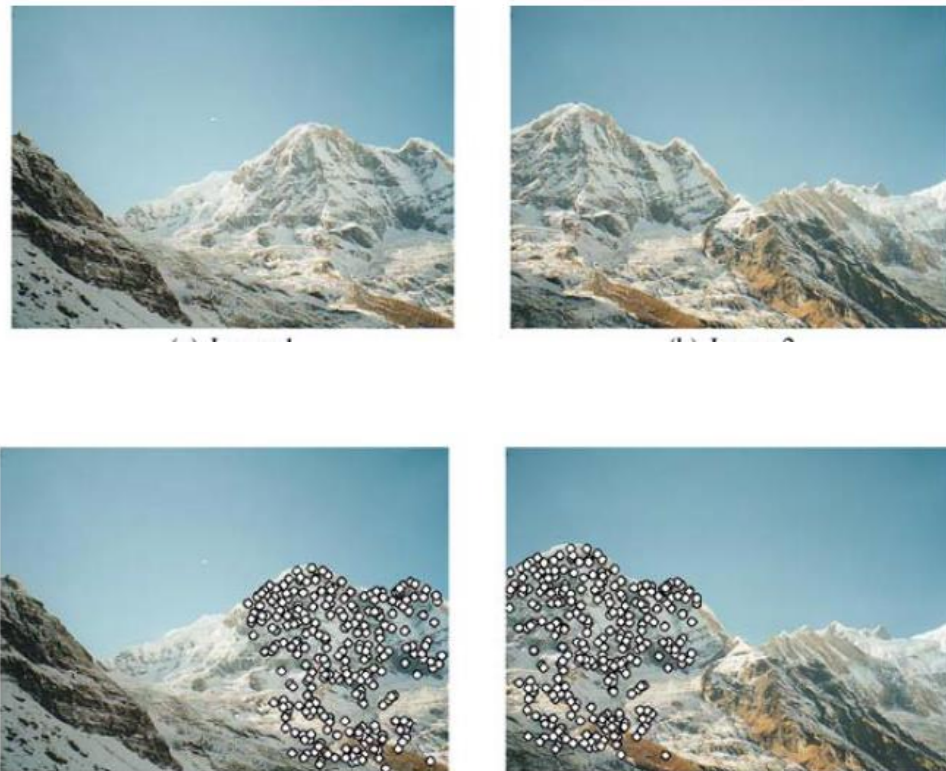
Εικόνα 4.1 Μοντέλο γεωμετρίας επιπόλων

Οπότε κάθε προβολική-χαρτογράφηση μεταξύ δύο οποιονδήποτε προβολών με το ίδιο κέντρο προβολής ονομάζεται ομογραφία(homography). Αναλύοντας την σχέση 4.1 βλέπουμε ότι το x_2 αντιστοιχίζεται στο διάνυσμα $\begin{pmatrix} wx^2 \\ wy^2 \\ w \end{pmatrix}$ και το x_1 σε $\begin{pmatrix} x \\ y \\ I \end{pmatrix}$.

Στην περίπτωση των πανοραμάτων επιλέγουμε σύνολα αντιστοιχίσεων των χαρακτηριστικών $r=4$ και υπολογίζουν στην συνέχεια την ομογραφία H μεταξύ τους χρησιμοποιώντας τον άμεσο γραμμικό μετασχηματισμό (direct linear transformation DLT) μία μέθοδο των Hartley και Zisserman [26]. Επαναλαμβάνουμε την διαδικασία για $n = 500$ δοκιμές και επιλέγουμε τη λύση η οποία έχει τον μέγιστο αριθμό inliers. Λαμβάνοντας υπόψη την πιθανότητα ότι μία αντιστοίχιση χαρακτηριστικού είναι σωστή μεταξύ ενός ζεύγους εικόνων (η πιθανότητα inlier) η οποία είναι p_i , η πιθανότητα να βρεθεί η σωστή μετατροπή μετά από n δοκιμές είναι:

$$p(H \text{ να είναι σωστή}) = 1 - (1 - p_i^r)^n \quad (4.2)$$

Ύστερα από ένα μεγάλο αριθμό δοκιμών η πιθανότητα να βρεθεί η σωστή ομογραφία είναι πολύ μεγάλη. Για παράδειγμα, για πιθανότητα inlier $p_i = 0.5$ η πιθανότητα να μην βρεθεί η σωστή ομογραφία ύστερα από 500 δοκιμές είναι 1×10^{-14} .



Εικόνα 4.2 Πάνω:Οι δύο αρχικές εικόνες ,Κάτω: Τα RANSAC inliers των αρχικών εικόνων

4.2.2 Πιθανοτικό μοντέλο για την επαλήθευση της αντιστοίχισης εικόνας

Για κάθε ζεύγος των αντιστοιχισμένων εικόνων έχουμε ένα σύνολο αντιστοιχισμένων χαρακτηριστικών τα οποία έχουν μια γεωμετρική συνοχή (RANSAC inliers) και ένα σύνολο χαρακτηριστικών τα οποία βρίσκονται στο εσωτερικό της περιοχής επικάλυψης (RANSAC outliers). Η ιδέα του μοντέλου επαλήθευσης είναι να συγκρίνει τις πιθανότητες του συνόλου inliers/outliers το οποίο δημιουργήθηκε από ένα μια σωστή αντιστοίχιση εικόνας η μία εσφαλμένη αντιστοίχιση.

Για μια δεδομένη εικόνα ο συνολικός αριθμός των χαρακτηριστικών στην περιοχή αλληλεπικάλυψης είναι n_f και ο αριθμός των inliers n_i . Το γεγονός ότι αυτή η εικόνα ταιριάζει σωστά ή λάθος αντιπροσωπεύεται από την δυαδική μεταβλητή $m \in \{0, 1\}$.

$$p(f^{(1:n_f)} | m = 1) = B(n_i : n_f, p_1) \quad (4.3)$$

$$p(f^{(1:n_f)} | m = 0) = B(n_i : n_f, p_0) \quad (4.4)$$

Όπου p_1 είναι η πιθανότητα ένα χαρακτηριστικό να είναι inlier και να δίνει μια σωστή αντιστοιχία και p_0 η πιθανότητα να είναι μια εσφαλμένη ατιστοιχία. Το σύνολο των μεταβλητών των αντιστοιχισμένων χαρακτηριστικών είναι $\{f^{(i)}, i = 1, 2, \dots, n_f\}$, ο αριθμός των inliers $n_i = \sum_{i=1}^{n_f} f^{(i)}$ και B η διωνυμική κατανομή (Binomial distribution).

$$B(x : n, p) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \quad (4.5)$$

Επιλέγουμε τιμές για p_1 και p_0 , 0.6 και 0.1 αντίστοιχα. Μπορούμε πλέον να υπολογίσουμε την πιθανότητα μια αντιστοίχιση να είναι σωστή χρησιμοποιώντας έναν κανόνα του Bay:

$$p(m=1 | f^{(1:n_f)}) = \frac{p(f^{(1:n_f)} | m=1)p(m=1)}{p(f^{(1:n_f)})} \quad (4.6)$$

Μόλις καθιερωθούν τα ζεύγη αντιστοίχισης μεταξύ των εικόνων, μπορούμε να βρούμε πανοραμικές ακολουθίες που συνδέονται από σύνολα που αντιστοιχίζουν εικόνες. Αυτό μας επιτρέπει να αναγνωρίζουμε πολλαπλά πανοράματα σε ένα σύνολο εικόνων, και να απορρίπτουμε εικόνες με θόρυβο οι οποίες δεν αντιστοιχίζονται με καμία άλλη εικόνα.

4.3 Συρραφή εικόνων

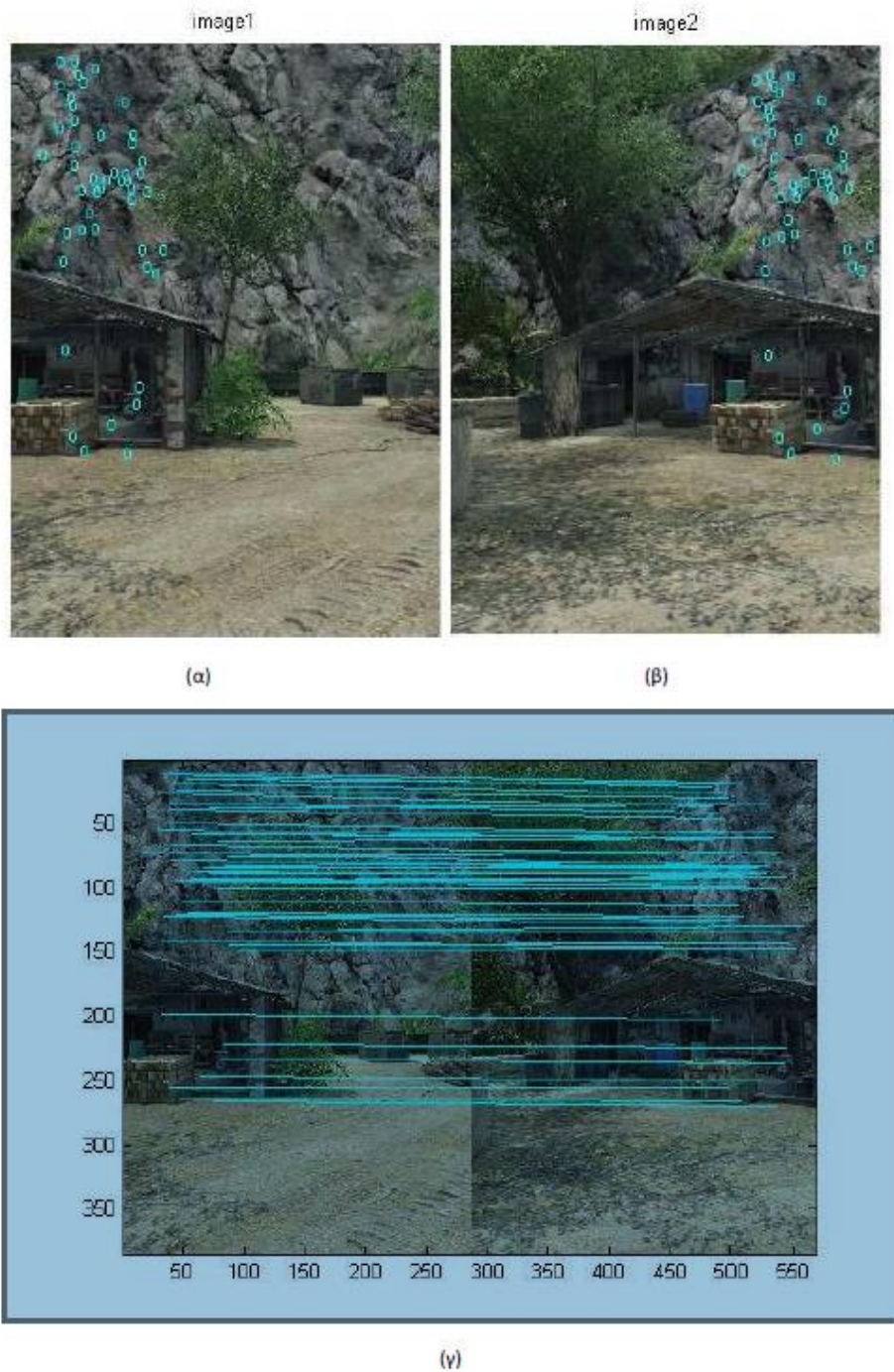
Μέχρι τώρα έχουμε βρει τα κοινά σημεία κλειδιά ανάμεσα στις δύο εικόνες, την αντιστοίχισή τους και το μοντέλο που τα συνδέει, δηλαδή την ομογραφία. Το επόμενο και τελευταίο βήμα είναι να ενώσουμε τις εικόνες. Χρησιμοποιώντας το μετασχηματισμό, την ομογραφία, και ορίζοντας τη μία από τις δύο σαν αρχική μετασχηματίζουμε την άλλη ώστε να την προσαρμόσουμε στην πρώτη. Αν όμως ενώσουμε απλώς έτσι τις εικόνες, θα είναι ορατή η ακμή της ένωσής τους. Γι' αυτό για την ανάμειξη των εικόνων θα χρησιμοποιήσουμε τη μέθοδο των Λαπλασιανών πυραμίδων.

Έστω ότι θέλουμε να ενώσουμε την εικόνα 1 με την εικόνα 2. Θα ορίσουμε σαν εικόνα αναφοράς τη μία από τις δύο (έστω την δύο) και χρησιμοποιώντας την ομογραφία θα προσαρμόσουμε την 1 πάνω στην 2. Αρχικά μέσω της ομογραφίας H θα βρούμε τις νέες συντεταγμένες των τεσσάρων γωνιών της εικόνας 1. Οι γωνιές αυτές είχαν συντεταγμένες: $(1,1)$, $(w,1)$, $(1,h)$ και (w,h) , όπου w το πλάτος της εικόνας 1 και h το ύψος της και με τη βοήθεια της ομογραφίας θα βρούμε ποιες θα είναι οι νέες τους συντεταγμένες σε σχέση με την εικόνα 2, που θα αποτελούν και πάλι τις γωνίες της στρεβλωμένης πλέον εικόνας.

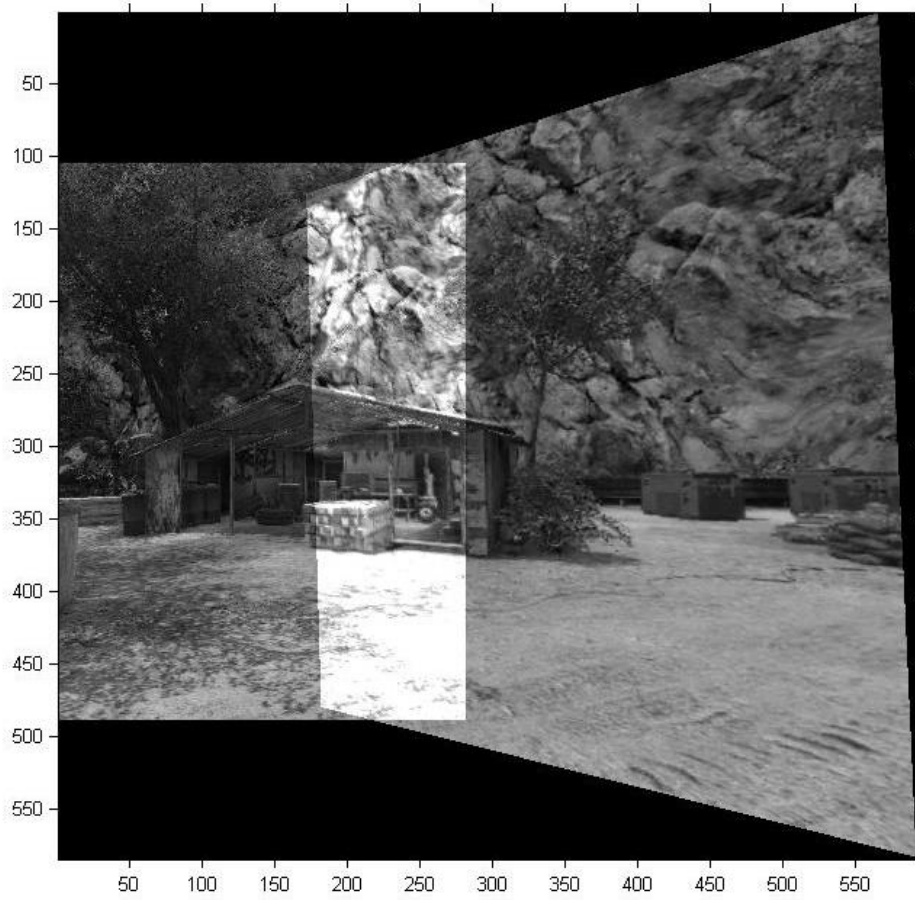
Αν οι νέες συντεταγμένες είναι (a,b) , (c,d) , (e,f) , (g,h) . Στη συνέχεια με τη βοήθεια της αντίστροφης ομογραφίας βρίσκουμε την στρεβλωμένη πλέον εικόνα 1 ($warpred$), χρησιμοποιώντας γραμμική παρεμβολή.

Αυτό το το οποίο γίνεται σε αυτό το τελικό στάδιο είναι να βρούμε κάθε σημείο της εικόνας 1 με ποιο σημείο της εικόνας 2 ταιριάζει.

Παρακάτω δίνεται ένα παράδειγμα όπου αρχικά εντοπίζονται τα σημεία τα οποία πρέπει να αντιστοιχιστούν στη συνέχεια συνδέονται και τελικώς γίνεται η πολυπλοκότερη συρραφή τους.



Εικόνα 4.3 Στην α και β εικόνα εντοπίζονται με τον SURF τα σημεία τα οποία θα αντιστοιχίσουμε. Στην γ τα σημεία ενώνονται



Εικόνα 4.4 Η εικόνα που προκύπτει από την ένωση των δύο παραπάνω εικόνων

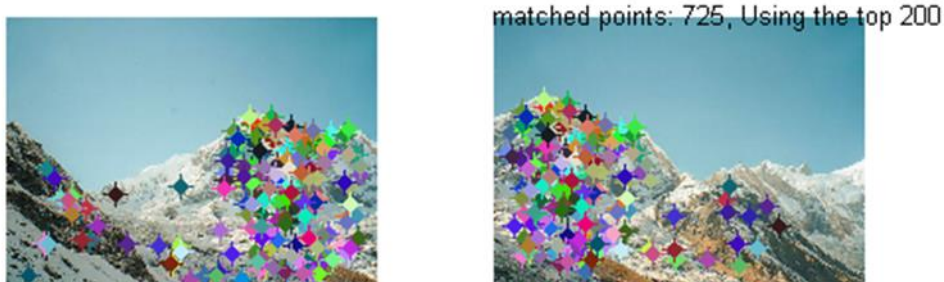
Κεφάλαιο 5.

Παραδείγματα – Αποτελέσματα

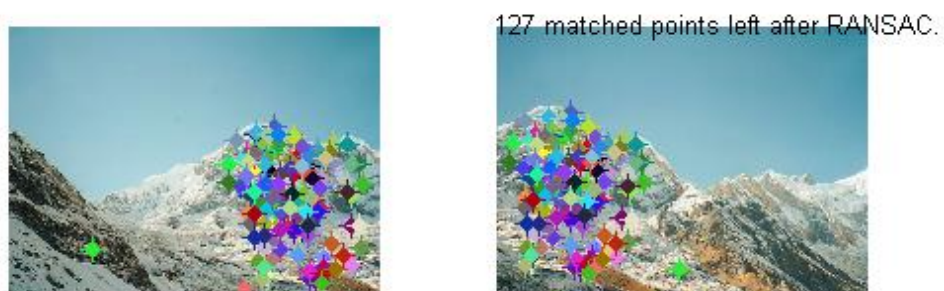
Σε αυτό το κεφάλαιο θα δοθούν τρία παραδείγματα της παραπάνω εφαρμογής. Στην εφαρμογή μας χρησιμοποιήσαμε τους αλγόριθμους SURF και RANSAC έτσι ώστε να έχουμε τα επιθυμητά αποτελέσματα. Ακόμη η εφαρμογή αναπτύχθηκε σε περιβάλλον MATLAB.

5.1 Πρώτο παράδειγμα

Σε αυτό όπως και όλα τα παραδείγματα δίνονται αρχικά δύο διαφορετικές εικόνες του ίδιου περιβάλλοντος. Στις εικόνες (5.1 και 5.2) φαίνονται οι αρχικές εικόνες ύστερα από τον εντοπισμό των σημείων ενδιαφέροντος (SURF) και το φιλτράρισμα αυτών των σημείων (RANSAC) αντίστοιχα.



Εικόνα 5.1 Το αποτέλεσμα που προκύπτει μετά την χρήση του SURF



Εικόνα 5.2 Το αποτέλεσμα που προκύπτει μετά τον υπολογισμό της ομογραφίας (RANSAC)

Τέλος παρουσιάζεται το τελικό πανόραμα που εξάγεται από την εφαρμογή μας στην εικόνα (5.3).



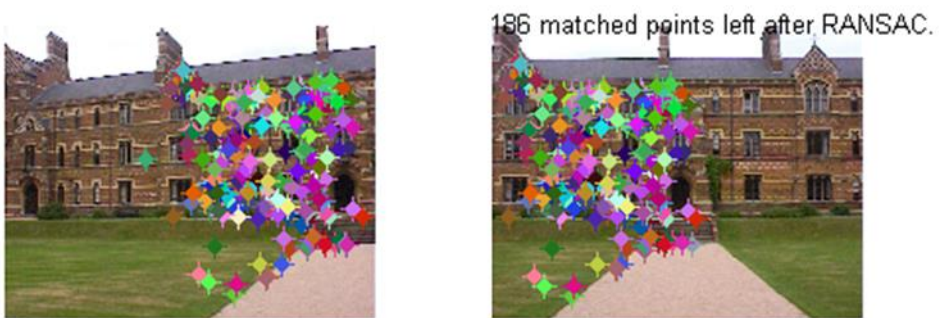
Εικόνα 5.3 Το τελικό πανόραμα από την ένωση των δύο αρχικών εικόνων

5.2 Δεύτερο Παράδειγμα

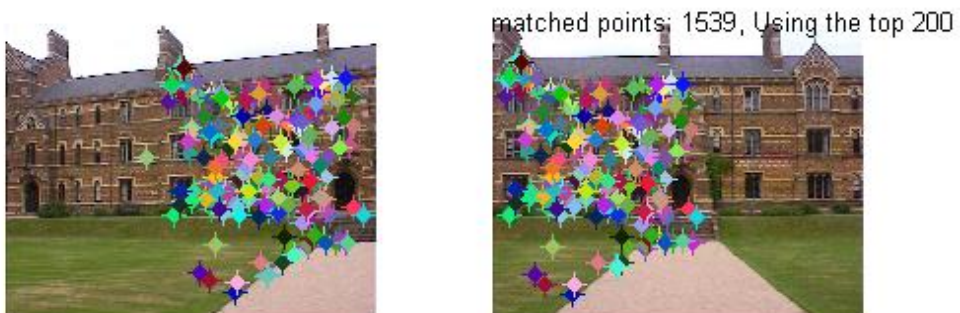
Και σε αυτό το παράδειγμα ακολουθείται η ίδια διαδικασία όπως παραπάνω. Αυτό που παρατηρείται είναι πως το τελικό πανόραμα που λαμβάνουμε είναι αρκετά πιο αποτελεσματικό από το προηγούμενο παράδειγμα, κάτι το οποίο οφείλεται στις αρχικές εικόνες και τον τρόπο με τον οποίο έχουν ληφθεί.



Εικόνα 5.4 Τελικό πανόραμα



Εικόνα 5.5 Αποτέλεσμα μετά την επεξεργασία του αλγορίθμου RANSAC



Εικόνα 5.6 Οι αρχικές εικόνες και τα σημεία ενδιαφέροντος τους (SURF)

Βιβλιογραφία

1. Rafael C. Gonzalez, Richard E. Woods. Digital Image Processing. s.l. : Addison-Wesley, 1992.
2. Starck, Jean-Luc. Image processing and data analysis: The multiscale approach.
3. Bijaoui, A. Wavelets and the analysis of astronomical objects,. s.l. : Large-Scale Structures in Nonlinear Physics, Springer-Verlag, 1991.
4. Algorithms de la transformation en ondelettes. s.l. : Ondelettes et Paquets d'Ondes, CoursCEA/EdF/INRIA, 1991.
5. Bijaoui. Wavelets and astronomical image analysis, Wavelets, Fractals and Fourier Transforms: New Developments and New Applications. Oxford, : M. Farge, J.C.R. Hunt and J.C. Vassilicos, Oxford University Press, 1993.
6. A. J. Ahumada, Jr., A. M. Rohaly, A. B. Watson. Image Discrimination Models Predict Object Detection in Natural Backgrounds. s.l. : Investigative Ophthalmology and Visual Science, 1995. Vol. 36.
7. P. Carbonetto, N. de Freitas, and K. Barnard. A statistical model for general contextual object recognition. s.l. : Proc. ECCV, 2004.
8. Koller, G. Heitz and D. Learning spatial context: Using stuff to find things. s.l. : Proc. ECCV, 2008.
9. Triggs, N. Dalal and B. Histograms of oriented gradients for human detection. s.l. : Proc CVPR, 2005.
10. S.Bileschi, L. Wolf and. A critical view of context. s.l. : IJVC, 2006.
11. Santosh K. Divvala, Derek Hoiem, James H. Hays, Alexei A. Efros, Martial Hebert. An Empirical Study of Context in Object Detection.

12. D. Hoiem, A. Efros, and M. Hebert. Recovering occlusion boundaries from a single image. ICCV : s.n., 2007.
13. Hua, Ming Yang and Ying Wu and Gang. Context-Aware Visual Tracking. s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.
14. Lowe, D.G. (2001) "Local feature view clustering for 3D object recognition", IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, 682-688
15. Lowe D. G. (2004) "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision 60, 2, 91-110
16. Bay H., Tuytelaars T., and Van Gool L. (2006) "Surf: Speeded up robust features", Proceedings of European Conference on Computer Vision, Graz, Austria
17. Martin A. Fischer, Robert C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, 1981
18. D. Capel, An Effective Bail-out Test for RANSAC Consensus Scoring, 2005
19. O. Chum, J. Matas, J. Kittler, Locally optimized RANSAC, 2003
- 20 P.H.S. Torr, P.A. Beardsley, D.W. Murray, Robot Vision, 1994
- 21 V. Nguyen, A. Martinelli, N. Tomatis, R. Siegwart, A Comparison of Line Extraction Algorithms using 2D laser Rangefinder for Indoor Mobile Robotics, 2005
- 22 K.A. Berthe, Y. Yan, S. Dembele, Image Detection in Real-time based on Fuzzy Fractal Theory, 2008
- 23 R.I. Hartley, In Defense of the Eight-Point Algorithm, 1997
- 24 R.I. Hartley, Minimizing Algebraic Distance, 1997
- 25 R.I. Hartley, Estimation of Relative Camera Positions for Uncalibrated Cameras, 1992
- 26 R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2003ment.
27. Yi Ma, Stefano Soatto, Jana Kosecka, Shankar S. Sastry, An Invitation to 3-D Vision: From Images to Geometric Models, Springer-Verlag New York Inc, 2004
28. P.H.S. Torr, D.W. Murray, The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix, 1997

29. Mikolajczyk K. and Schmid C (2005) "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis & Machine Intelligence, 27, (10), 1615–1630